CONSTRUCTION OF COMPREHENSIVE ACHIEVEMENT EXAMINATIONS FOR NAVY OFFICER CANDIDATE PROGRAMS

Astia

Prepared under auspices of

BUREAU OF NAVAL PERSONNEL

PERSONNEL ANALYSIS DIVISION

TRAINING RESEARCH BRANCH

under contract with

OFFICE OF NAVAL RESEARCH

NOVEMBER 1953

AMERICAN INSTITUTE FOR RESEARCH

PITTSBURGH 32, PENNSYLVANIA

CONSTRUCTION OF COMPREHENSIVE ACHIEVEMENT EXAMINATIONS FOR NAVY OFFICER CANDIDATE PROGRAMS

Prepared Under the Auspices of

Bureau of Naval Personnel, Personnel Analysis Division, Training Research Branch

Under Contract With

Office of Naval Research

John B. Carroll, Project Director
Benjamin Schohan, Research Associate

Project Designation No. NR 154-138

November 1953

American Institute for Research
Pittsburgh, Pennsylvania

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (continued)

ABSTRACT

1. The purpose of this project was to construct a new series of examination materials for measuring the end-of-course achievement of students in naval officer candidate programs. The examinations were to test "integrated knowledge" rather than mere memory for factual information. For the purposes of the project, an examination was defined as measuring "integrated knowledge" if it incorporated a substantial proportion of items measuring problem-solving and decision-making ability. These items would presumably draw upon a variety of knowledges and at the same time test the ability to apply these knowledges in realistic situations.

2. On the basis of the above definition of an "integrated examination," a new set of examinations in naval science, to be administered as year-end achievement tests to NROTC students in May 1953, was prepared. These examinations, while incorporating some materials which had been used in the examinations of previous years, included a substantial proportion of new items judged to measure problem-solving and decision-making ability. In the course of preparing these examinations, the examinations in use as final achievement tests at the Officer Candidate School, Newport, R. I., were extensively revised. All examinations were constructed to parallel as closely as possible the respective curricula of the NROTC and OCS programs.

3. In statistical analyses designed to evaluate the examinations constructed, the following conclusions concerning the OCS tests were reached:

(a) The time-limits which had been set were appropriate for measuring power rather than speed.

(b) The tests met reasonably satisfactory standards of reliability and score-dispersion. Further improvement in test reliability would, however, be desirable.

(c) The subject-matter covered by the tests was in general very heterogeneous.

(d) The scores for the various department examinations were highly correlated, suggesting a single general factor of achievement in Officer Candidate School.

(e) The items of the tests showed a reasonably satisfactory distribution of item difficulty.

(f) A reasonably large number of items showed satisfactory validity against a criterion which consisted of the total weighted scores on all examinations.

(g) High validities were found equally often for items appearing to measure sheer factual information and for items appearing to require problem-solving and decision-making.

It should be pointed out that no truly satisfactory criterion (based, for example, upon observations of shipboard performance) was available for evaluating the new examination materials which were constructed.

4. An effort was made to sort out the items of the OCS examinations in terms of different varieties of achievement. The Wherry-Gaylord iterative factor-analysis procedure, applied to 72 representative items, yielded five linearly independent factors. This analysis was later expanded to include all 410 items of the examinations. One factor was interpreted as a residual error factor. The four factors which remained were rather highly correlated, a result which suggested that their isolation and separation was chiefly due to the capitalization of chance independencies. There was certainly no strong evidence for the separation of different varieties of achievement in OCS. On the contrary, the chief conclusion was that achievement in OCS is quite uniform; that is, individuals tend to perform at the same level of achievement in all aspects of the curriculum. Furthermore, individuals tend to perform at the same level of achievement regardless of whether

they are tested with items covering factual information or with items appearing to measure problem-solving and decision-making ability. Nevertheless, to the extent that the factors identified by the Wherry-Gaylord procedure could be interpreted, the following types of achievement seem to exist with at least some degree of independence:

I. Verbal reasoning and judgment (Key E100).

II. Spatio-temporal reasoning (Keys E200 and E300, which were quite highly correlated).

III. Ability to remember factual information (Key E400).

5. The technique of "individual operating characteristics" was applied to the interpretation of the factor keys. This technique is designed to present the probability with which an individual getting any given score will pass items of any given difficulty. The technique thus enables the interpretation of test scores in terms of actual performance on items rather than merely in terms of norms. Complete data on the operating characteristics of OCS factor scores are presented in the body of this report. The graphical presentations of operating characteristic curves confirm the observation made elsewhere that the content of the examinations is extremely heterogeneous.

6. An attempt was made to compare groups of OCS and NROTC students with respect to performance on naval science achievement tests. When representative groups of OCS and NROTC students were compared on material common to the two curricula and freshly reviewed by both groups, the NROTC students were on the average superior to the OCS students. This was particularly true of 3rd and 4th year NROTC students; this result suggests the possible operation of a selective factor throughout the four-year NROTC curriculum.

However, when OCS students were compared with 4th year NROTC students on material covered in the first three years of the NROTC course and hence not

freshly reviewed by the NROTC students, the OCS students were clearly superior, particularly so after appropriate statistical controls were imposed. Thus, even though 4th year NROTC students are superior to OCS students on material covered in the 4th-year NROTC curriculum, they show considerable lack of retention for the material covered in earlier years.

RECOMMENDATIONS FOR FURTHER RESEARCH

1. In the construction of new examination materials, the approach utilized in the present project should be further explored. Items identified in this report as appearing to measure problem-solving and decision-making ability might be used as models for the construction of new examination items. Also, this approach could be extended into certain areas of the naval science curriculum which were not thoroughly worked on in the present project, such as Engineering and Damage Control, and Military Justice. Descriptions of realistic situations which would provide the necessary context for the construction of items measuring problem-solving and decision-making abilities might be developed.

2. It would be desirable to validate the special factor keys developed in the present research against on-the-job shipboard performance. They may also be used as tentative criteria for new items which may be constructed, and for this purpose the items composing these factor keys should continue to be included in future examinations. At the same time, further research could be undertaken to examine the factorial composition of naval science examinations in order to see whether further evidence can be found for differentiating factual-knowledge examinations from what have been called "integrated" examinations in the present research.

FOREWORD

This report was prepared by the American Institute for Research in its Cambridge, Massachusetts office under Contract Nonr-890(02), Project NR154-138. The project was initiated by, and administered under the direction of the Bureau of Naval Personnel, Personnel Analysis Division, Training Research Branch, with Dr. Eugene D. Carstater and Mr. Leo J. Brogan acting as project monitors; and under contract with the Office of Naval Research, Psychological Services Division, Personnel and Training Branch.

Acknowledgement is hereby made of the generous cooperation extended to the project by the following Naval officers in making members of their respective staffs available as subject-matter consultants and/or providing test scores.

Captain R. S. Johnson, USN, Commanding Officer, Naval Schools Command, NTS, Newport, Rhode Island.

Captain D. V. Gladding, USN, Professor of Naval Science, NROTC Unit Harvard University.

Captain H. B. Southworth, USN, Professor of Naval Science, NROTC Unit, Tufts College.

Captain L. M. Smith, USNR, Commanding Officer, Volunteer Research Unit No. 1-1, Boston, Massachusetts.

LCDR R. L. Collins, USNR, Commanding Officer, Volunteer Harbor Defense Unit No. 1-1, Boston, Massachusetts.

Instructors of Naval Science at the Officers Candidate School, Newport, Rhode Island, and at Tufts College and Harvard University, in addition to Naval Officers serving in local USN Reserve units, aided greatly by voluntarily serving as subject-matter consultants in the test construction phase of the project.

Marvin W. Herrick of the project staff was directly responsible for establishing the test item files and for aiding in the construction of items used in several of the examinations.

D. David Bourland, Jr. and Mrs. Ruth Twombly, also of the project staff, served as most conscientious and facile statistical clerks.

Thanks are due Mr. John Alman, of Boston University, for his assistance in connection with certain IBM tabulations. Other IBM tabulations were prepared in the statistical department o the Pittsburgh office of the American Institute for Research.

Credit must also be given to Professor P. J. Rulon of Harvard University and Dr. Norman Frederiksen of the Educational Testing Service, Princeton, New Jersey, for having served as consultants at one stage of the project. Their advice was of great usefulness.

INTRODUCTION

This project was initiated in an attempt to provide an improved examination program to be used by the Navy in evaluating the achievement of students enrolled in schools for officer candidates.

Before the inception of the project, the Bureau of Naval Personnel had expressed a desire to improve the examinations then in use. The Bureau felt that a new series of examinations designed to test the candidate's "integrated knowledge" of the subject-matter of Naval Science courses would provide a more valid means of assessing performance and of assigning lineal precedence to graduates, as required by law.

In addition, the Bureau of Naval Personnel had expressed a desire to obtain a means of establishing objective comparisons among trainees in the three types of general line programs, the Officer Candidate School (OCS), the Naval Reserve Officer Training Programs (NROTC), and the Reserve Officer Candidate Program (ROC). These programs differ considerably with espect to purpose, type, and duration of training, and recruiting of trainees:

(a) The Officer Candidate School at the U.S. Naval Base, Newport, R. I. conducts an 18-week course emphasizing "practical," vocational aspects of all phases of basic naval science. Students are college graduates.

(b) The Naval Reserve Officer Training Program is conducted at 52 colleges and universities, in conjunction with the regular academic courses at these institutions. The course lasts four years, and covers both "practical" and "theoretical" aspects of naval science, with an emphasis on the latter.

(c) The Reserve Officer Candidate Program is conducted at one or more naval installations during the summer. The trainees are college students and others who, since they are not attending colleges having NROTC Units, find

it possible to take this training only during the summer. The curriculum is similar to that established for OCS, but may depart from it in some respects, if only because of the fact that the program is conducted at naval installations other than Newport and hence with a different instructional staff.

In view of the differences among these programs, the Bureau of Naval Personnel has not felt it wise to prescribe identical final examinations for them. Hence, the problem of comparing these programs in some objective manner has appeared resistant to any simple solution.

In order to solve the problems posed by the above-cited needs of the Navy, it was believed that certain new procedures of test analysis would have to be developed and applied.

Accordingly, this project established three basic objectives:

1. The modification of existing final achievement examinations for naval officer candidates to meet more adequately the requirements set forth by the research sponsor, namely, that the examinations should measure the extent to which candidates could demonstrate an "integrated" knowledge of naval science subjects rather than merely memory for factual details. The examinations would be designed to measure the ability to solve specific problems requiring technical and professional skill, and the ability to make sound decisions in novel situations.
2. The development and application of a technique for comparing, with respect to performance at the end of the course, naval officer candidates enrolled in the three general line programs described above: NROTC, OCS, and ROC.
3. The accomplishment, so far as possible, of basic research into methods of investigating the content homogeneity of achievement examinations and methods of scaling and interpreting test performance.

At the outset, the work of the project was planned in two phases. Phase I involved the construction of a series of comprehensive examinations in naval science, and thus was designed to fulfill research objective (1) above. Phase II was to be concerned with the detailed analysis of data, and hence with research objectives (2) and (3). Phase II was to be undertaken only after Phase I was nearly complete.

This report is divided into two parts, corresponding to these phases of work.

PART ONE

THE CONSTRUCTION OF ACHIEVEMENT EXAMINATIONS IN NAVAL SCIENCE

CHAPTER I. DEFINING EXAMINATION OBJECTIVES

The Examination and the Curriculum

> ". . . the function of ROTC is to implant in a prospective officer the knowledge of his craft which he must have if he is to function as an officer when he goes on active duty. It reminds us that any specialized profession calls for memorizing a body of exact facts -- medical students, for example, have to learn anatomy, and they are apt not to like it in contrast to the more speculative subjects they studied in college. How well a young officer has mastered the details of his training may spell the difference between survival and death for himself and the men under his command. An infantry officer must be able to work a gun; whether he understands the chemistry of explosives is purely incidental to his responsibilities as a platoon leader or company commander. To know what buttons to push on an automatic calculator is more important to a young antiaircraft officer than to comprehend the mathematical theories that made it possible. Admittedly this sort of knowledge or skill is not a normal element of a college education; but this is not to say that as a field of study military science must live in a narrow isolation from the academic curriculum. Total war is more than a strictly military problem. The 'know why' is an essential element of the 'know how', and should be part of the equipment of an ROTC graudate."

So writes President Dodds of Princeton, in presenting the case for professional military education in civilian colleges and universities, in a recent article in the Atlantic Monthly (March, 1953).

The conflicting demands of fact and theory in professional education have never been adjudicated to the complete satisfaction of everyone, and probably never will be. On the one hand, there always exist, in any field of professional training, more than enough facts to fill a curriculum to bursting, and for each of these facts one can find experts who will say that its acquisition is an absolute necessity on the part of the trainee. On the other hand, there are those who will say that a deep knowledge of theory will suffice to get any man by in any situation.

In the formation of a curriculum, the conflict between fact and theory can often be bypassed, on the ground that the conflict can be compromised by teaching fact and theory in equal measure, but in the design of an examination to cover this curriculum the conflict must be faced openly. An examination is a device for evaluating the achievement of the student in terms of the behaviors he is expected to have acquired as a result of training. What kind of behavior is desired in a naval officer candidate at the end of his training? This question, in the broad setting we have just described, is one which had to be faced in the conduct of the present project.

The sponsor of the project, the Bureau of Naval Personnel specified that the examinations which were to be constructed for the evaluation of Naval Officer Candidate graduates should be truly "comprehensive and integrated," not in the sense of a rigorous coverage of the subject-matter, but in terms of a better sampling of the actual skills and knowledges demanded of the NROTC graduates when assigned to the fleet. It suggested that the tasks set in the examinations should be concerned with the ability to make decisions in novel situations and the ability to solve problems, and that at the same time the solution of these tasks should call for a wide range of factual knowledge, so that the tests would measure the extent to which examinees had integrated their knowledge in the whole context of the material presented in the curriculum. It is thus seen that the Bureau of Naval Personnel took a middle position with respect to the relative roles of theory, fact, and practical application.

The project staff sought to arrive at a more specific, operationally applicable definition of "comprehensiveness" and "integration" with reference to the actual conditions surrounding the project. The conditions of which we speak had to do with the facilities available for tryout of the examinations and the present state of instructional programs where testing was to be conducted.

Specifically, it was seen necessary to suit the examinations to actual, current curricula rather than to the more nearly ideal curricula which might be established at some time in the future. Furthermore, since the examination materials which were eventually to be used in connection with the NROTC program had to be developed in the context of instruction at the Officer Candidate School at Newport, R. I., a continual view of what was practicable in the light of differences between these curricula had to be maintained.

Three ways of defining the concept of a "comprehensive, integrated examination" in naval science were considered, and although all were in some measure useful in guiding efforts in test construction, only one of them was found to be realistic as a primary point of departure. The three definitions are listed below, with comments on their relevance to possible kinds of items in the situation outlined above.

1. "Integrated" items could be conceived as those which test organized combinations of subject-matter from different fields. For example, one could set a problem which involved facts and principles from two fields, such as <u>Weapons</u> and <u>Operations</u>*.

This definition was rejected because the NROTC curriculum is organized in such a manner that subject fields are taught over a period of four years. It would therefore be impossible to construct a truly comprehensive test covering several subject fields for <u>each</u> year of Naval science. An alternative would be to construct cumulative achievement examinations for administration at the end of each year, but such tests would be essentially tests of retention and would require NROTC units to provide cumulative reviews. However, this definition might have been workable if tests were being prepared solely for the Officer Candidate School where all the subject fields are taught concurrently.

* The OCS curriculum is organized in six departments: Orientation and Military Justice, Naval Weapons, Operations, Seamanship, Navigation, and Engineering.

2. "Integrated" items could be conceived as those which test combinations of knowledges required in the duties performed as an Ensign. Thus, the items would in effect constitute "verbalized performance" items. Each item would ask the examinee to state how he would perform in a certain situation and it would require an integrated knowledge of several facts in order to arrive at a correct statement of proper performance.

This definition was rejected in principle for two reasons:

a. If valid items were to be constructed on the basis of combinations of knowledges required of a newly commissioned Ensign in the performance of his duties aboard ship, detailed job analyses of these duties and requirements would have to be available. It must be remembered that the duties and responsibilities of Ensigns differ greatly depending upon the types of vessel to which they are assigned. Thus the responsibilities of an Ensign on a battleship are not identical to those of an Ensign aboard an LST. Therefore, several job analyses would be required, corresponding to the duties of Ensigns aboard different types of naval craft. Such job analyses were not available to project personnel.

b. Even if the requisite types of job analyses were available, their use in the construction of achievement examinations for officer candidates would be questionable in the light of present NROTC and OCS curricula. Training in the officer candidate programs (line) is not specific to certain ships. The curriculum intent is to acquaint the trainees with the Navy and with naval procedures and to provide them with a broad background of knowledges and requirements connected with the operation, maintenance, and armament of naval vessels in general. Nevertheless, newly-commissioned Ensigns graduating from OCS are not expected to be

able to perform shipboard duties except at the lowest level of responsibility. It is only after they receive several months of shipboard, on-the-job training in specific duties that they are entrusted with independent responsibility. Thus, examinations based on the duties of fleet Ensigns might be in conflict with the intent and design of officer candidate program curricula.

3. Lastly, "integrated" items might be conceived as those which incorporate decision-making or problem-solving functions by the integration of knowledge; i.e., rather than testing on a specific knowledge "A," each item would require the examinee to manipulate knowledges "A," "B," and "C" in order to arrive at the correct answer "D." In addition, the item would require the examinee to perceive relations and implications existing between various facts. These "knowledges" represent items of instruction to which the candidate has been exposed during the course of the officer candidate training program.

This last definition was accepted to guide the work of construction in the test items.

It was assumed that items conforming to definition 3, and also in some degree to definition 2, could best be constructed by outlining typical sequences of activities in which a new Ensign might be engaged, and building items based upon these reaslistic situations. Such situations usually involve a complex of factors to be taken into consideration in solving problems. Hence, the examinations are expected to be predictive of how well examinees will be able to solve problems arising in their shipboard on-the-job training and in their later naval careers.

The extent to which it is possible to test "integrated" knowledge is partly a function of the form of the examination. The above considerations relating to ways of measuring "integrated knowledge" apply largely to examinations of the multiple-choice type. At a conference of project consultants held on 14 December 1952,

various alternative means of examining were proposed. For example, Dr. Norman Frederiksen suggested that examinees be shown a sound film on which they would take notes. This sound film would present shipboard situations in a realistic fashion. At appropriate points, the examinee would be asked to make certain decisions. After the correct answer was given, the film would proceed to other situations, to be handled in a similar fashion. Dr. Frederiksen also pointed out that the multiple-choice format reduces the opportunity for the individual to demonstrate creative problem-solving ability. It would perhaps be feasible, he suggested, to have examinees write solutions to problems in essay form, in duplicate. To adapt this procedure for objective scoring, one would have the examinee turn in one copy of his essay answer, then code the copy he retains in terms of multiple-choice options. Finally, both Dr. Frederiksen and Dr. Rulon suggested that work-sample or performance tests be constructed as a further means of testing "integrated knowledge." Unfortunately, due to project deadlines (which called for the construction of the NROTC examinations by 15 February 1953), none of these suggestions could be followed in the actual work of the project.

It should be emphasized that the NAVPERS Officer Candidate Achievement Examinations are not used as a basis for disqualification. When necessary, trainees are either disenrolled or "washed back" prior to the stage at which the achievement examinations are administered. Grades on these examinations do, however, comprise one of several criteria employed by the Navy to establish the lineal precedence of commissioned officers within their respective school classes. Thus, there seems some basis for the belief that examinees are fairly well motivated. Project personnel gained the impression that motivation in taking the examinations is especially high at NROTC units.

CHAPTER II. TEST CONSTRUCTION PROCEDURES

At the Officer Candidate School (Navy), the curriculum covers six areas of instruction: Orientation and Military Justice, Operations, Seamanship, Navigation, Naval Weapons, and Engineering and Damage Control. Instruction at OCS is given concurrently in the six areas during the 15 week course.

The NROTC program consists of training given during four college years in certain areas as follows:*

(Freshmen) Naval Science I - Orientation, Seamanship, Operations

(Sophomores) Naval Science II - Naval Weapons, Operations

(Juniors) Naval Science III - Navigation, Seamanship, Operations

(Seniors) Naval Science IV - Military Justice, Engineering, Seamanship.

It might appear from the above that the NROTC and OCS curricula are practically identical. Appearances, however, are deceptive. Although overlap existed between the two curricula in some areas of instruction, there were large differences in others. In effect, examinations had to be separately planned and constructed for use in OCS and in NROTC.

At the outset, project personnel were provided with copies of NAVPERS achievement examinations administered to NROTC midshipmen at the end of each of the four years of training. These sets of examinations were dated May 1951 and May 1952. NAVPERS achievement examinations administered in May 1952 in the six OCS areas of instruction were also provided. Copies of quarterly examinations and quizzes

---

* This breakdown with respect to areas of instruction at NROTC is an impressionistic summary of the curriculum as prescribed every year in an official BuPe publication (e.g., U.S. Navy Standard Curriculum for NROTC (line), NAVPERS 91828). The NROTC program is categorized by the navy simply in terms of years, e.g. NS-1, NS-2, rather than by subject material.

administered to trainees at the Officer Candidate School were also made available. Finally, item analyses of the 1951 and 1952 NROTC NAVPERS achievement examinations were furnished to aid in the item construction task.

A. Construction of Achievement Examinations for Officer Candidates (OCS)

The initial task of the project was to revise the OCS May 1952 NAVPERS achievement examinations in the six subject areas for administration to officer candidates at Newport just prior to graduation.

First, the 1952 OCS examinations and the May, 1952 NROTC examinations were subjected to detailed inspection for two purposes: (a) to gain an impression of the types of items previously in use, and (b) to evaluate the subject-matter coverage of existing examinations. This inspection of previous examinations was carried out by project personnel in collaboration with appropriate instructional personnel at the Newport OCS. For the purpose of appraising item-types, items were classified in three categories: "problem-solving," "factual-informational," and "decision-making." An attempt was made to introspect into the mental processes which examinees would have to employ in order to answer the items correctly. Consideration was given to the nature of the instruction which had been received; for example, an item which might appear to an outsider a "problem-solving" item might be judged merely "factual-informational" because the item called for a set solution presented in the classroom as something to be memorized. While it is recognized that mental functions overlap, an attempt was made to classify items in terms of the chief or crucial element involved.

As was expected, the large majority of items on all examinations were of the informational type, as may be seen from Table 2.1. The single exception was in the navigation examinations, which contained actual navigational problems; hence the items were classified as "problem-solving." It was seen, in any case, that a guiding principle in revising the tests would be an attempt to provide

more items of the "problem-solving" and "decision-making" types. Nevertheless, it was recognized that revised examinations should contain a reasonable proportion of "factual-informational" items.

TABLE 2.1

NUMBERS OF ITEMS CLASSIFIED IN EACH OF THREE CATEGORIES
May 1952 OCS and NROTC Examinations

		FACTUAL-INFORMATIONAL		PROBLEM-SOLVING		DECISION-MAKING		TOTAL
		No.	%	No.	%	No.	%	No.
OCS	Orientation and Military Justice	62	82.7	0	0	13	17.3	75
OCS	Operations	55	73.3	17	22.7	3	4.0	75
OCS	Seamanship	32	71.1	3	6.7	10	22.2	45
OCS	Navigation	24	40.0	36	60.0	0	0	60
OCS	Naval Weapons	71	94.7	4	5.3	0	0	75
OCS	Engineering and Damage Control	71	94.7	4	5.3	0	0	75
NROTC	First Year	145	96.7	2	1.3	3	2.0	150
NROTC	Second Year	137	91.3	11	7.3	2	1.3	150
NROTC	Third Year Part I	55	73.3	13	17.3	7	9.4	75
NROTC	Third Year Part II	0	0	30	100.0	0	0	30
NROTC	Fourth Year	134	89.3	5	3.3	11	7.3	150

With respect to coverage of the curriculum, it appeared that the existing examinations represented a balanced sampling of material; that is, they emphasized various portions of the curriculum roughly in proportion to the time spent in teaching these portions.

Upon completion of this analysis of existing examinations, test item files were established for each of the six areas of OCS instruction.

Next, actual construction of test items was undertaken for the final examinations to be administered to OCS Class VII in October 1952. The limited objective established at that time was to modify existing NAVPERS OCS examinations so that (a) the content would be more closely aligned with the curriculum currently in use or, with planned changes in the curriculum and (b) only items regarded as useful in measuring overall competence of the officer candidate would be retained. Although many factual items would be retained, effort was to be expended in selecting or constructing items measuring a more integrated type of knowledge with the emphasis on problem solving, situational items.

Test items in the six examinations were editorially improved, completely revised or rewritten according to the following procedure:

1. Based upon the curriculum and study guides, a new tentative test outline was developed for each of the six OCS examinations.

2. For each examination the backlog of items was assembled in accordance with the new test outline. In those cases where a number of items existed for a specific outline topic, the items which appeared "best" in terms of content importance and measurement characteristics were selected by project personnel. Item revisions were suggested directly on the item cards.

3. At this point, the new test outlines and the selected items and reviews were discussed with OCS instructor personnel. In collaboration with these instructors, each item selected by project personnel, together with the suggested revisions was reviewed. Further revisions were made where necessary. If an item was found to be totally unsuitable, the instructors were asked to inspect the remaining items in the file for a possible substitution or modification. If no suitable substitute item was available, a new item was constructed. New

items were also constructed to cover areas of the test outline for which no items previously existed. Some items required the use of pictorial material, which was prepared at Newport by Navy artists.

4. The final examination items were edited and assembled for typing and reproduction. The placements of correct answers within the five alternatives were randomized.

5. In order to insure accuracy, OCS officials reviewed the final reproduction copy.

The reproduction copy for the six examinations to be administered to Class VII was delivered to the Academic Director, OCS, Newport, R.I. along with three IBM score keys for each test.

Subsequent to the administration of the examinations to OCS Class VII, item analyses of the six examinations were performed, using Flanagan's procedure involving groups of 9, 20, 42, 20, and 9 per cent.* The average grade achieved on the six achievement examinations constituted the criterion against which all items were validated. This uniform criterion was used for all items for the following reasons: (1) It was desired to select items for use in future examinations which would measure "integrated knowledge," insofar as possible. The average grade on all six examinations was believed to represent the best criterion available for this "integrated knowledge." (2) The use of a uniform criterion would effect considerable savings in the time required to prepare the data for analysis. The alternative procedure would have been to use the total score on each examination as the criterion against which items in that examination would be evaluated. Such a procedure would have tended to select items uniquely measuring different subject-matters rather than to select items measuring an

---

* Flanagan, J.C. The effectiveness of short methods for calculating correlation coefficients. Psychol. Bull., 1952, 49, 342-348.

integrated knowledge of the whole course. The actual results of item analysis will be presented and discussed in a later section of this report.

In the light of item analysis data, the examinations in Seamanship, Weapons and Navigation were further revised for administration to OCS Class VIII, December 1952 graduation. The examinations in Orientation and Military Justice, Operations, and Engineering and Damage Control constructed for OCS Class VII (October, 1952) were again administered to OCS Class VIII, with only very minor revisions.

Of the six new examinations thus far constructed, the Naval Weapons was most completely revised. This examination was now considered to be directly aligned with the OCS curriculum. For example, in former examinations several questions were devoted to atomic theory, principles, and weapons. These were eliminated from the new examination since atomic weapons and theory were no longer taught at OCS. The examination items are for the most part "integrated" items; many are in situational contexts. That is, they do not test on specific details, but require the integration of knowledges acquired throughout the course. Use was made of the device of presenting a realistic developing situation and then asking a number of questions relating to the situation.

The Navigation examination (part 2) has always contained a celestial problem requiring a plot of a "day's work." This problem represents an integration of all knowledges and procedures required with respect to navigating a vessel at sea from one point to another. The problem set for the OCS Class VIII was revised in accordance with the 1952 Nautical Almanac. Prior examinations were based on the 1950 almanac.

The Seamanship examination for Class VIII was revised to include more items on Rules of the Road; otherwise, it underwent no great change.

Again, the six examinations administered to OCS Class VIII were subjected to an item analysis.

Subsequently, the Operations examination was completely revised for administration to OCS Class IX (March 1953 graduation). With the revision of this examination, work was completed on test construction for OCS.

Since the construction of the March 1953 OCS Operations examination represented one of the project's most thoroughgoing attempts to prepare an "integrated" examination of the desired type, details on its preparation may be of interest.

In December, 1952, the project director spent a full day discussing plans for this examination with Operations instructors at Newport. He asked them to envisage a typical tour of duty which a newly-commissioned officer might take, and to specify the kinds of experiences which they might have on such a tour of duty. After considerable probing, the investigator elicited a description of experiences which a new ensign might typically have on board a destroyer as it moves from a Navy yard, through a shakedown cruise, and out to fleet operations in a combat area. For example, the new Ensign is first assigned as Custodian of Registered Publications with collateral duties as Chief Watch Officer. The subject-matter experts then helped the investigator draft a number of item-ideas based on this situation. Questions were drafted relating to the chain-of-command structure under which the ensign would operate, the problems involved in establishing a file of registered publications, the guarding of radio channels, the preparation, evaluation, and distribution of messages, etc. In a later phase of his tour of duty, the ensign is envisaged as being assigned to a watch as Officer of the Day (OOD). Item-ideas relating to this situation involved various problems of ship-maneuvering, communications, CIC, anti-submarine warfare, etc.

Inspection of the items resulting from this mode of attack shows that many of them must frankly be regarded as measuring mere factual information. For one thing, having relevant factual information is doubtless crucial in many of the

situations on which these items are based. Secondly, in an examination it is hardly possible to present situations in all the detailed context in which they would occur in real life. Thirdly, instructors at Newport insist that the curriculum at OCS is of such a nature that they cannot give students any real understanding or impression of the way things happen in shipboard operations; a student can get the "feel" of being a naval officer only after he has had experience in the fleet. Therefore, any questions which might probe into the subtler and more complicated aspects of fleet activities would be almost meaningless to officer candidates. Nevertheless, to the extent that the items which appear in the final form of the March 1953 Operations examination are embedded in somewhat realistic contexts, they may cause the examinee to answer not merely in terms of bookish information but in terms of his understanding, however dim, of the way in which this information is relevant in actual fleet operations.

Thus, the Officer Candidate School, Newport, was provided with new achievement examinations for each of its six areas of instruction. These examinations have now been reproduced in quantity by OCS Newport and are designed to be administered to future graduating officer candidate classes for some time to come, i.e., until new examinations are prepared. The total number of items in each examination is as follows:

Orientation and Military Justice....................75 items

Operations...50 items

Seamanship...50 items

Navigation...60 items

Naval Weapons......................................75 items

Engineering and Damage Control.....................75 items

B. Construction of Achievement Examinations for Midshipmen (NROTC)

It was established that the NROTC 1952-53 and OCS 1952 curricula were not completely comparable with respect to the subject-matter taught, a situation which made the construction of common examinations extremely difficult, if not impossible. NROTC trainees are given a more extensive course involving relatively more theory, whereas OCS trainees receive a course which emphasizes practical aspects. Moreover, many items of instruction presented to NROTC trainees are omitted altogether at OCS.

To illustrate: the OCS curriculum is subdivided into six areas of instruction, as already explained. The NROTC curriculum is subdivided into four college years of instruction, each of the years cutting across the six OCS areas of instruction. In the first-year NROTC program, 39 class periods are devoted to the study of Naval History, but in the Orientation course at OCS, Naval History is not formally taught, the trainees being required to read outside of class a text on Naval History. Thus, great emphasis is placed on this subject in NROTC whereas it appears to be only incidental to the OCS curriculum. In fact, questions on naval history have been completely eliminated from the OCS Orientation and Military Justice examination. Conversely, certain subject-matter in Operations such as CIC and communications is more heavily emphasized at OCS than in NROTC training, where there is considerable tendency to avoid teaching information based on classified documents.

The celestial problem in the OCS Navigation examination was constructed to conform with the 1952 Nautical Almanac in use at OCS. However, it was necessary to reconstruct the whole problem for the NROTC examination because midshipmen were to be provided with the 1953 Almanac.

The construction of identical examinations for OCS and NROTC in Naval Weapons was impossible. The two curricula are not comparable, at least from a test

construction point of view. In fact, great difficulty was experienced at the outset by project personnel in constructing items for incorporation in both examinations. The opinions of NROTC subject-matter instructors differed greatly from those of OCS consultants on which items were valid and fair. Items suggested by NROTC consultants were vigorously opposed by OCS consultants and vice versa. The primary basis for disagreement was difference in the two curricula. Certain materials presented in the NROTC program are not taught at OCS, for example, radar in connection with gunnery, and atomic theory.

Thus, it was evident that examinations constructed for OCS use would be inadequate for the NROTC program, both in length (total number of items per examination) and scope. It was therefore decided to assemble new examinations for the NROTC program.

Essentially the same procedure as that employed in assembling the OCS examinations was followed in constructing examinations for NROTC. Test outlines based on the 1952-53 curriculum were designed for each of the four Naval Science (NROTC) examinations. NROTC instructors at Harvard University and Tufts College were consulted for their views on the outlines specifically with respect to curriculum coverage and areas of emphasis. The test outlines were modified in accordance with their suggestions pertaining to areas of emphasis wherever necessary.

On the basis of item analysis data, valid and applicable items from the newly constructed OCS examinations and former NROTC examinations were selected and/or revised for inclusion in the NROTC examinations. Additional items were constructed in collaboration with Navy subject-matter consultants wherever necessary to complete the coverage indicated by the test outlines. Items to be incorporated in each of the four new NROTC examinations were then reviewed by NROTC instructor personnel. Further item revisions were made as necessary. Throughout this procedure, as before, an attempt was made to select or to

construct items of the desired "integrated" type, but it cannot be said that this attempt was completely successful in view of the large amounts of informational material which had to be covered. For example, little could be done to alter the general nature of the items testing Naval History in the First Year NROTC examination.

The items for each of the four achievement examinations, together with pertinent item analysis data, were forwarded to the Bureau of Naval Personnel for final review prior to the typing of reproduction copy.

The total number of items for each of the Officer Candidate Achievement Examinations (NROTC) is as follows:

Naval Science I (Freshmen)............................150 items

Naval Science II (Sophomores)..........................150 items

Naval Science III (Juniors)............................105 items

Part I - 75 items
Part 2 --30 items

Naval Science IV (Seniors).............................150 items

Final reproduction copy of the Officer Candidate Achievement Examinations (NROTC) and an Examiners Manual (modeled on those provided for previous years) for the administration and scoring of the examinations were submitted to the Bureau of Naval Personnel, Washington, D.C. for reproduction and distribution to the 52 college NROTC units. Prepunched scoring keys for the examinations were also supplied to BuPers in sufficient quantity for direct distribution to the NROTC units.

Summary and Discussion of Phase I of the Project

Phase I was concerned chiefly with the provision of a new set of examinations to be administered at NROTC units in May, 1953. Certain desiderata for these examinations were specified by the project sponsor, chiefly to the effect that the examinations should test "integrated" knowledge rather than mere knowledge of

isolated facts. To the extent that it was possible to do so in view of the target dates for completion, this objective was achieved. A new set of examinations for NROTC was provided, and considerable progress was made in developing items calling for problem-solving, decision-making, and weighing of factors in complex situations. In the course of preparing examinations for NROTC, a new set of examinations for OCS was prepared.

It is pertinent to comment on the extent to which the approach described above was found practicable. This will be done by considering each of several areas of instructions.

Nearly all the subjects covered under Naval Operations, including tactics, maneuvering board, communications, CIC, and ASW, are amenable to the construction of problem-solving and decision-making items. The project made considerable progress in this direction, as may be seen by inspection of the March, 1953 OCS Operations examination (NAVPERS 18307, March 1953, available at OCS Newport).

However, it was found impossible to include "integrated" type items on Operations in the NROTC examination, since the subject-matter in this area taught in NROTC is of a relatively elementary character. The only items of the problem-solving type included in NROTC examinations were maneuvering board problems (see, for example, questions 61 - 70 in the Third Year NROTC Examination, Part I, NAVPERS 18288, May 1953).

Of the subjects taught under Orientation and Military Justice, those having to do with military insignia, naval regulations, naval customs, the naval establishment, naval administration and correspondence, and naval history were in general found not amenable to the construction of problem-solving or decision-making items. Most of the items of knowledge taught in this area exist, at least for beginning trainees, as isolated bits of information and cannot be thrown into any underlying rationale or structure. While it is conceivable that problem-

solving situations involving these facts could be constructed as examination items, the effort required and the space needed on the examination paper could hardly be justified. A more conventional, direct approach has been deemed preferable in testing these subject-matters. Subject-matter in naval law and naval leadership, however, lends itself more readily to the problem-solving approach. Naval law includes much factual information which must be presented and tested as such, but at the same time presents possibilities for constructing items which depict some of the types of situations which a naval officer might commonly encounter. Attempts in this direction are represented by items 110-113, 121, and 125 on the Fourth Year NROTC Examination (NAVPERS 18290, May 1953). However, the situations cannot be so complex that they present technicalities which can be solved only by an officer with special legal training.

Likewise, the subject-matter of Naval Leadership lends itself to items of the situational type. An especially good example is provided by item 132 of the above-cited Fourth Year NROTC examination (NAVPERS 18290, May 1953). This item was especially constructed by project personnel to test knowledge of certain principles of leadership and ability to weight various factors in the situation; to our knowledge, no situation exactly like this was discussed in NROTC classrooms. (This does not preclude the possibility of "coaching" in future years, of course.) Unfortunately, the validity of this item is unknown since it was included only in the May 1953 NROTC examination, given too late to be item-analyzed by this project.

The subject of <u>Navigation</u> is obviously amenable to testing by problem-solving items. In fact, a major section in previous examinations had been constructed in this form, and it was only necessary for this project to adapt this section to use with several successive editions of the Nautical Almanac. The reader will recall that this section (Part II, items 31-60) consists of 30 items based upon a "day's work" in which the examinee is required to solve typical navigational problems.

There remain, of course, many aspects of the subject-matter of navigation which do not lend themselves conveniently to problem-solving items. For example, questions regarding the meaning of map symbols, the classification of buoys, navigational publications, and the like can be best handled by direct questions designed to measure factual information. Such questions are found among the first thirty items on the OCS Navigation examination (NAVPERS 18309).

A similar situation exists in connection with the subject matter of <u>Seamanship</u>. Items on ship nomenclature, ships of the Navy, and many aspects of deck seamanship (e.g., "What is the length of a standard shot of anchor chain?") could conceivably be set in "integrated, problem-solving" form, but they are more easily constructed as single, direct multiple-choice items. On the other hand, topics such as weather, the duties of the OOD, rules of the road, and the like lend themselves to problem situations. The project made some progress in constructing such items, but more could be done along these lines.

As has been mentioned above, a major effort of the project was directed towards the construction of "integrated" items in Naval Weapons. The approach was to present typical situations, and then ask questions relating to decisions required in these situations. In the opinion of project personnel, the examination in Naval Weapons now in use at OCS (NAVPERS 18310) represents a close approximation to the best that can be done in the light of the present curriculum. For example, items 10-13 therein present a situation in which the examinee is to assume that he is "assigned to the main battery of a Baltimore class heavy cruiser which is engaged in a night surface action with enemy heavy cruisers." The examinee is required to indicate the correct decisions in the following cases:

(item 10) which type of projectile to use to inflict maximum below-deck damage;

(item 11) which radar spotting method to use;

(item 12) how the rangekeeper should be adjusted for differing generated range and observed range;

(item 13) how battery control should be shifted after the forward controlling director suffers a casualty.

Nevertheless, even in situational-type items it was found necessary to resort to questions of a more factual nature, in order to cover certain theoretical aspects of the curriculum. For example, item 8 in NAVPERS 18310 begins with the following lead: "The depth charge attack fails. Your ship therefore launches a hedgehog attack using a 7.2" projector charge. This charge derives its propulsive force from . . ." The question thus becomes quite factual and specific. It would probably be undesirable to eliminate such questions altogether.

The subject-matter of Engineering and Damage Control was not adequately investigated by the project from a test-construction point of view, due to lack of time and personnel. The final forms of examinations covering this subject cannot be said to contain any large number of items of a problem-solving type. The impression was gained, however, that much could be done in the direction of constructing tests of integrated knowledge in this area. This is true despite the fact that the curricula of both OCS and NROTC require the presentation of much theoretical knowledge concerning such matters as the physics of pressure, heat, and gravity; characteristics of boilers, engines, and the like. It is envisaged that in further test-construction efforts in this area, items could be presented in the context of typical engine-room and damage-control happenings. Questions could be asked about measures to be taken in various types of emergencies. Questions about ship-loading also would lend themselves to the use of the problem-solving approach.

During the test-construction phase of this project, the staff found itself constantly searching for tangible ways by which one could be assured that examination questions of the desired "integrated" type were being constructed.

Although it was possible to identify items which the staff could agree upon as appearing to be of the desired type, subjective agreement was not a scientifically satisfying criterion. Phase I of the project was thus concluded with only a hope and a promise: somewhere in the wide variety of items which had been constructed, items of the desired type could be reasonably expected to exist. It was believed, however, that these items could be objectively identified only by resort to statistical factor-analytic procedures. Such procedures would presumably separate several types of items, one or more of which could be appraised as items of an "integrated" type, others of which could be considered as more "factual" in nature. For this reason, at least half the total time of the project was devoted to statistical analysis, now to be described in Part II of this report.

PART TWO

EXPERIMENTAL AND STATISTICAL ANALYSES

CHAPTER III. GENERAL STATISTICAL EVALUATION OF EXAMINATIONS

Introduction

Before proceeding to more advanced phases of statistical analysis, the project staff desired to appraise the extent to which the examinations constructed in Phase I met the usual criteria of good tests--appropriate forms of score distribution, high reliabilities, and the like,--and also to appraise in a preliminary way whether the various tests, organized chiefly in terms of subject-matter, measured different aspects of achievement.

All the analyses reported in this chapter are based on examinations prepared for and administered to various classes of officer candidates at the U. S. Naval School (Officer Candidate), Newport, Rhode Island. In the case of the examinations prepared for administration to NROTC students in May 1953, data could not be obtained soon enough for analysis before the expiration of the contract. Since the NROTC examinations are generally similar in form to the OCS exams, it is probable that certain generalizations about the OCS examinations can be extended to apply to NROTC examinations.

Test Score Distributions

One of the first things which had to be determined was whether the time-limits provided for the examinations were adequate to allow nearly all examinees to finish trying all items. We espouse the point of view that officer candidate achievement tests should measure accuracy of knowledge rather than speed in answering test items. Except in certain situations (e.g., solving maneuvering board problems under combat conditions), speed in solving problems is probably not as important as accuracy in successful performance of naval officers, and certainly not in the training of naval officer candidates.

Accordingly, distributions were made of the item-number of the last item attempted, for the six final examinations given to Class VI at OCS (this class graduated July 1952). The results are shown in Table 3.1, based on 10 randomly selected sections of the class. In view of the large proportions of examinees completing all items, it was concluded that the previously established time-limits were adequate. Accordingly, no major changes in the length or time-limit of the OCS tests were made in subsequent revisions, except that the time-limit for the Seamanship examination was increased to 40 minutes in view of its lengthening to 50 items. It should also be pointed out that the March, 1953 Operations examination contained only 50 items instead of the previous 75. No change in time-limit was made, however, because the items were chiefly of the situational problem type and were expected to require more time per item.

The form of the test-score distributions was next considered. (The score used in all cases was the number right.*) When plotted on arithmetical probability paper, all score distributions were seen to be close approximations to normal distributions. It was not considered worthwhile to apply exact statistical tests of normality. In view of the approximate normality of distributions and the fact that the means and standard deviations are presented in Table 3.2, it is also not considered worthwhile or necessary to present here the graphs showing the exact form of each distribution.

Test Means, Standard Deviations, and Reliabilities

The data to be discussed in this section are presented in Table 3.2, showing various test statistics for each of the numerous examination forms administered to the classes VI to IX at OCS Newport. Test forms in each subject are indicated by A, B, and C, together with primes; use of letters B or C indicates major

* The score in the Navigation Examination was, however, arrived at by the Newport-established formula which weighted Part I by a factor of 2 and Part II by a factor of 3.

TABLE 3.1

DATA ON NUMBER OF ITEMS ATTEMPTED IN SIX OCS EXAMINATIONS, CLASS VI

Test	Time-limit (minutes)	No. of Items	Distribution of "last item attempted"				N
Orientation and Military Justice	60	75	All examinees completed 75 items.				287
			#	f	#	f	
Operations	60	75	75	272	70	1	283
			74	1	69	0	
			73	1	68	1	
			72	2	67	1	
			71	2	66	2	
Seamanship	36	45	All examinees completed 45 items.				282
Navigation Part I	100	30	All examinees completed 30 items				254
			#	f	#	f	
Navigation Part II		30	30	223	24	5	254
			29	2	23	2	
			28	1	22	3	
			27	12	21	0	
			26	3	20	0	
			25	1	19	1	
					18	1	
			#	f			
Naval Weapons	60	75	75	285			286
			74	0			
			73	0			
			72	1			
Engineering and Damage Control	60	75	All examinees completed 75 items.				281

TABLE 3.2

MEANS, STANDARD DEVIATIONS, AND RELIABILITY AND HOMOGENEITY DATA

FOR OCS EXAMINATIONS

Note: All data on reliability and homogeneity are based on random samples of 200 cases used for item analysis.

Class and Date Graduated	N	Test Form	Max. Score	$\bar{X}$	σ	K-R (20) rel.	rel. (1 item)	H_t
<u>Engineering</u> <u>and</u> <u>Damage</u> <u>Control</u>								
VI (July 1952)	681	A	75	54.5	6.9	*	*	*
VII (Oct. 1952)	684	A'	75	50.0	7.0	.706	.031	.049
VIII (Dec. 1952)	772	A'	75	48.8	7.4	.743	.039	.059
IX (March 1953)	834	A'	75	48.3	7.1	*	*	*
<u>Navigation</u> (Part I only)								
VI (July 1952)	681	A	30	20.2	3.1	*	*	*
VII (Oct. 1952)	684	B	30	22.1**	3.1**	.513	.034	.063
VIII (Dec. 1952)	772	B'	30	20.0**	3.0**	.499	.032	.061
IX (March 1953)	834	B'	30	*	*	*	*	*
<u>Navigation</u> (Part II only)								
VI (July 1952)	681	A	30	22.7	3.8	*	*	*
VII (Oct. 1952)	684	B	30	23.5**	3.8**	.782	.107	.155
VIII (Dec. 1952)	772	B'	30	22.7**	3.3**	.696	.071	.184
IX (March 1953)	834	B'	30	*	*	*	*	*
<u>Navigation</u> (Parts I and II combined, with part I weighted by 2 and part II weighted by 3.)								
VI (July 1952)	681	A	150	*	*	*	*	*
VII (Oct. 1952)	684	B	150	114.7	14.0	.770+	*	*
VIII (Dec. 1952)	772	B'	150	112.2	13.3	.715+	*	*
IX (March 1953)	834	B'	150	114.7	13.9	*	*	*
<u>Operations</u>								
VI (July 1952)	681	A	75	47.4	6.0	*	*	*
VII (Oct. 1952)	684	B	75	48.7	6.4	.683	.028	.048
VIII (Dec. 1952)	772	B	75	48.2	6.2	.669	.026	.048
IX (March 1953)	834	C	50	31.7	5.2	.632	.033	.048

(Table continued on next page)

TABLE 3.2 (continued)

Class and Date Graduated	N	Test Form	Max. Score	$\bar{X}$	σ	K-R (20) rel.	rel. (1 item)	H_t
<u>Orientation and Military Justice</u>								
VI (July 1952)	681	A	75	51.0	5.1	*	*	*
VII (Oct. 1952)	684	B	75	52.0	5.2	.533	.015	.027
VIII (Dec. 1952)	772	B	75	51.7	5.2	.546	.016	.029
IX (March 1953)	834	B	75	51.6	5.1	*	*	*
<u>Seamanship</u>								
VI (July 1952)	681	A	45	34.2	3.3	*	*	*
VII (Oct. 1952)	684	B	50	39.6	3.4	.419	.014	.028
VIII (Dec. 1952)	772	B'	50	35.0	4.3	.541	.023	.041
IX (March 1953)	834	B'	50	35.8	4.0	*	*	*
<u>Naval Weapons</u>								
VI (July 1952)	681	A	75	54.5	6.5	*	*	*
VII (Oct. 1952)	684	B	75	52.0	6.2	.683	.028	.051
VIII (Dec. 1952)	772	C	75	47.7	7.3	.746	.038	.064
IX (March 1953)	834	C	75	45.2	7.1	*	*	*

* Not computed.

** Based on the random sample of 200 cases used for item analysis.

\+ Estimated by Gulliksen's formula (74) in Chapter 20 (Gulliksen, H. Theory of Mental Tests, New York: Wiley, 1950). The correlation between Part I and Part II of the Navigation examination was .27 in Class VII; this figure was used as an estimate of the correlation in Class VIII, which was not actually computed.

revisions from Form A, while a prime indicates a minor revision. For example, the <u>Seamanship</u> examination Class VII represented a considerable revision (Form B), but underwent only minor revisions thereafter (Form B'). Allusions will also be made to Table 3.3 which concerns the statistics of academic department grades at Newport.

With regard to test means and standard deviations, the following interpretations may be made:

(a) The test mean is anywhere from 60% to 78% of the total possible score. This is mentioned chiefly because of the practice at Newport of converting the test scores to the Navy 4.0 system of grading by the formula

$$\text{Navy grade} = 4\left(\frac{\text{Score}}{\text{Max. Score}}\right)$$

with the intention of obtaining an average Navy Grade of about 3.0 to 3.2. If the tests are to be used with this procedure in mind, the test mean should be from 75 to 80 per cent of the total possible score. Nevertheless, the Navy formula has the disadvantage of assuming that tests measure in terms of absolute standards. The characteristics of a test score distribution should be evaluated from the standpoint of relative rather than absolute standards. If tests of maximum reliability are desired, one guiding principle is that the test mean should be about in the center of the range from a pure chance score to the maximum score, i.e., about 60 per cent of the total possible score for tests with 5-choice items. In view of this, the OCS examinations reported on in Table 3.2 are slightly too easy. It should be remembered that the ratio between test mean and the maximum score is directly related to the average item difficulty. In a later section of this report, the distributions of item difficulties will be discussed.

(b) There is some tendency for the test means (at least in terms of average item difficulties) to decrease as one moves from Class VI to Class IX. This is

also reflected in the mean Bureau Examination averages reported in Table 3.3 which decrease from 2.841 in Class VII to 2.706 in Class IX (the difference being significant far beyond the .001 level). This result may be due to decreasing quality in the successive classes or to the failure of the examinations to keep pace with gradual changes in the curriculum. For example, the mean test score in the Engineering and Damage Control examination, which was subjected to only very minor revisions by this project, decreased from 54.5 to 48.3 from Class VI to IX. (This difference is also significant far beyond the .001 level.) On the other hand, the decrease is at least partly due to the fact that in successive revisions of the various examinations the project staff sought to avoid very easy items (particularly when they showed low validities), and in general to reduce the average item difficulty index (proportion passing). It also appears that when attempts were made to subject tests to major revisions (to the extent of constructing numerous new items of a situational problem type), the average item difficulty and hence the test mean tended to be reduced. This tendency seems to be illustrated in the case of the Naval Weapons and the Operations examinations. (It would have been possible to test the significance of the decreases only by use of analysis of covariance with suitable control variables such as Officer Candidate Battery scores, but this was not considered worth the necessary effort. Furthermore, this procedure would be rather unusual in using different dependent variables for the various classes.)

(c) The test standard deviations are such that $\pm 3\sigma$ from the mean covers anywhere from 51 to 78 per cent of the range from a pure chance score to a perfect score. It was thought desirable to increase these ranges with successive test revisions by appropriate selection of items, and there is some evidence that the project was successful in this regard, for example in the Operations and Naval Weapons examinations. Further progress in this direction could in theory be made by selecting items with a wider range of difficulty and with somewhat higher validities.

(d) Reliabilities of many of the examinations were computed from item-analysis data by means of the Kuder-Richardson formula (20).* Since the July 1952 examinations and most of the March 1953 examinations (the single exception being Operations, which was a completely new examination) were not subjected to item analysis, no reliability data for these are reported in Table 3.2.

The reliabilities of the examinations were disappointingly low, in spite of persistent efforts to select or to construct items which would be valid when judged against internal consistency criteria. It is of interest to note that one of the tests least worked on by the project remained one of the most reliable, namely, the Engineering and Damage Control examination, which had reliabilities of .706 and .743, respectively, in Classes VII and VIII. Taking the reliabilities of the latest examinations constructed by the project, one may rank the tests as follows:

	Reliability
Naval Weapons (Form C)	.746
Engineering (Form A')	.743
Navigation, Part II (Form B')	.696
Operations (Form C)	.632
Orientation and Military Justice (Form B)	.546
Seamanship (Form B')	.541
Navigation, Part I (Form B')	.499

* This is presented as formula (11) in Chapter 16 of Harold Gulliksen's Theory of Mental Tests (N.Y., Wiley, 1950). This reliability coefficient has often been regarded as a lower bound to the equivalent-form reliability. Lee Cronbach, (Cronbach, Lee S., "Coefficient alpha and the internal structure of tests." Psychometrika 1951, 16, 297-334.) however, has recently shown that it is also the average stepped-up reliability obtainable from all possible splits of a test into two parts. In any case, the formula seems practicable and pertinent to achievement examinations of the type studied here.

This, however, does not quite give the whole story. In the first place, the reliability of the composite score on the Navigation examination (Form B') may safely be estimated at .715 and thus is almost as reliable as the Engineering and Naval Weapons examinations. In the second place, the reliability figures given above are for differing test lengths. Accordingly, the reliability per item (given in the next to last column of Table 3.2) was computed for each test by applying the Spearman-Brown formula "in reverse."* In these terms, the tests

---

* In effect, this procedure estimates the reliability of a test of a standard length, namely one item. If r_x is the reliability of a test of n items, the reliability of one item (r_1) is, by the Spearman-Brown formula,

$$r_1 = \frac{\frac{1}{n} r_x}{1 + \left(\frac{1}{n} - 1\right) r_x} = \frac{r_x}{n - (n - 1) r_x}$$

It can be shown that this formula is very simply related to the expression presented by Harold Gulliksen as "a function of test reliability that is invariant with respect to changes in test length" (formula 22, Chapter 8, Theory of Mental Tests, N. Y., Wiley, 1950). Gulliksen's expression (in our notation) is:

$$n \left(\frac{1}{r_x} - 1\right)$$

and this is equal to $(1 - r_1) / r_1$. The use of r_1 as an invariant function of reliability is preferable to Gulliksen's expression because it is in the metric of correlation coefficient, whereas Gulliksen's expres ion ranges from .00 (for perfectly reliable tests) to $+\infty$ (for tests of zero reliability).

rank as follows (data are again for the latest test constructed):

	Reliability (1 item)
Navigation Part II (Form B')	.071
Engineering (Form A')	.039
Naval Weapons (Form C)	.038
Operations (Form C)	.033
Navigation Part I (Form B')	.032
Seamanship (Form B')	.023
Orientation and Military Justice (Form B)	.016

It appears that the reliability per item is a function of the extent to which the test measures technical information involving knowledge of mathematics and physics. However, it is not necessarily related to the "problem-solving" nature of the item, since the Engineering examination, with a relatively high per-item reliability, contained only a small number of such items. The tests showing lowest per-item reliabilities were those concerned with isolated bits of non-technical information related to the conventions, regulations, and customs of the navy and seamanship. Thus, there seems to be something inherent in the characteristics of various types of test content which is a factor in the maximum per-item reliabilities which can be obtained in testing these contents. Since per-item reliability estimates are also to some extent a function of average item-intercorrelation, the per-item reliability data can also be regarded as reflecting the content homogeneity within each of the several examinations.

In view of the relatively low reliabilities of the separate department examinations, one may reasonably be concerned with the reliability of the composite score computed at Newport and labeled "Bureau Examination average." This score is computed (at least, before converting to the Navy grading system)

by adding the raw scores on the examinations, with a weight of 1 for all examinations except Navigation (Part I) and Naval Weapons, which receive a weight of 2, and Navigation (Part II), which receives a weight of 3. The reliability of the composite score can then be estimated by an application of the correlation-of-sums formula. As shown by Gulliksen (formula 74 in Chapter 20, Theory of Mental Tests), this may be written

$$\text{rel.}_{(\text{wtd. composite})} = \frac{WCW'}{WRW'} ,$$

where W is a row vector of weights, C is the matrix of intercorrelations with test reliabilities in the diagonal, and R is the matrix of intercorrelations with unities in the diagonal. This formula was applied to the data for the Class VII and Class VIII examinations, using appropriate intercorrelations from Table 3.5. Since the formula assumes that the variables to be weighted are in standardized form, the row vector of weights was taken as the standard deviations of the weighted scores which were summed to make the composite. Data for the two parts of the Navigation examination were combined; thus, the weight for the Navigation examination was the actual standard deviation of its own composite score. (A footnote to Table 3.2 explains the one approximation that was involved in this procedure.) By this means, the reliabilities of the Bureau examination average were estimated to be as follows (standard errors of measurement in terms of the Navy grading system are also given):

		Reliability	$\sigma_{\text{meas.}}$
Class VII	(October 1952)	.888	.082
lass VIII	(December 1952)	.894	.085

In view of the total length of the battery, these reliabilities are not as high as might be desired, and need to be improved in further test construction efforts; likewise, the standard errors of measurement are uncomfortably large for accurate

ranking of students. It is probable that the composite could be made slightly more reliable by a different weighting of the tests; the optimal weights could be determined by the procedure outlined by Gulliksen in Section 12, Chapter 20, Theory of Mental Tests. This procedure is computationally complex, however, and would not produce a large enough increase in reliability to justify its use except as a final refinement. A more fundamental solution to the problem of greater reliability of the composite is to increase the reliability of the subtests.

It should be remarked, however, that the present research has been concerned more with the validity of the examinations than with their reliability. Thus, while the reliabilities of the successively modified tests do not show an upward trend, their subject-matter validity may have increased. Indeed, it is possible that "integrated" examinations will be found to have inherently low reliabilities. As has been pointed out, however, no external criterion against which to measure validity has yet offered itself.

(e) Another approach to content homogeneity is represented by Loevinger's coefficient of homogeneity, H_t.* This coefficient measures (on a scale from .00 to 1.00) the extent to which the variance of a test approaches the variance of what Loevinger calls a "perfectly homogenous test" in which all tetrachoric correlations between items would be unity. Such a test would also be perfectly scalable in Louis Guttman's sense. The writer's experience has been that as applied to most tests Loevinger's coefficient has very small values, and this is confirmed in the present instance. Ranking the latest forms of the examinations with respect to H_t, we get the following list:

* Loevinger, Jane. A systematic approach to the construction and evaluation of tests of ability. Psychol. Monog., 1947, 61, Whole No. 285.

	H_t
Navigation Part II	.184
Naval Weapons	.064
Navigation Part I	.061
Engineering and Damage Control	.059
Operations	.048
Seamanship	.041
Orientation and Military Justice	.029

The above ranking is highly related ($\rho = .82$) to the ranking with respect to per-item reliability; hence the two measures appear to be measuring much the same thing, or at least to be affected by the same test characteristics. The relatively high value obtained for Navigation, Part II, is consistent with the observation that this whole test is concerned with a single navigation problem.

It is of interest to note that in nearly every case, later forms of the tests tended to have higher homogeneity coefficients than earlier forms, even when the reliability coefficients did not increase. If these increases are significant (no statistical tests of significance are available, however), they may be due to the fact that the later revisions contained items of higher average validity.

<u>Means and Standard Deviations of Newport Academic Grades</u>

The means and standard deviations of academic grades, by department, for Classes VII, VIII, and IX are presented in Table 3.3. Data on the Bureau Examination Averages for Classes VI - IX are also presented. All these grades are in terms of the Navy grading system, which uses a scale running from 0.0 to 4.0. The Bureau Examination Average has a nominal weight of one-tenth in the Final Average Academic Grade. It was of interest to determine the actual weights of the various components in the Final Average Academic Grade. This was done by multiple-regression

TABLE 3.3

MEANS AND STANDARD DEVIATIONS OF ACADEMIC GRADES BY DEPARTMENT AND THE AVERAGE OF BUREAU EXAMINATIONS

ote: All grades represented here are in terms of the Navy system. (Scale from 0.0 to 4.0)

Class and Graduation Date	N	$\bar{X}$	σ
Final Average Academic Grade			
VII (Oct. 1952)	684	3.089	.182
VIII (Dec. 1952)	772	3.046	.210
IX (March 1953)	834	3.041	.197
Engineering Department Grades			
VII (Oct. 1952)	684	3.064	.217
VIII (Dec. 1952)	772	3.051	.276
IX (March 1953)	834	3.063	.232
Navigation Department Grades			
VII (Oct. 1952)	684	3.196	.281
VIII (Dec. 1952)	772	3.078	.347
IX (March 1953)	834	3.102	.294
Operations Department Grades			
VII (Oct. 1952)	684	3.153	.222
VIII (Dec. 1952)	772	3.175	.245
IX (March 1953)	834	3.078	.236
Orientation and Military Justice Grades			
VII (Oct. 1952)	684	3.111	.159
VIII (Dec. 1952)	772	3.151	.181
IX (March 1953)	834	3.139	.183
Seamanship Department Grades			
VII (Oct. 1952)	684	3.090	.186
VIII (Dec. 1952)	772	3.063	.288
IX (March 1953)	834	3.061	.208
Naval Weapons Department Grades			
VII (Oct. 1952)	684	3.030	.238
VIII (Dec. 1952)	772	3.022	.258
IX (March 1953)	834	3.034	.247
Bureau Examination Average (on 4.0 system)			
VI (July 1952)	681	2.840	*
VII (Oct. 1952)	684	2.841	.246
VIII (Dec. 1952)	772	2.718	.262
IX (March 1953)	834	2.706	.273

* Not computed

techniques for Class VIII, which was believed to be a representative sample for this purpose. Results are shown in Table 3.4, which shows that the relative weight of the Bureau examination average was 11.9%. It is conceivable that the Bureau examinations deserve somewhat more weight than this. Possibly there is hesitation to put more weight on the Bureau examinations because so doing would further depress the mean Final Average Academic Grade. Nevertheless, if some sort of standard score conversion formula were used for the Bureau examinations, instead of the Navy grading system, this could be avoided. Valid achievement examinations do not necessarily have a mean equivalent to 3.0 or thereabouts on the Navy grading scale.

Because the intercorrelations shown in Table 3.5 among the department grades, which like the Bureau Examination average, are in terms of the Navy grading system do not have a value of unity, it is not unexpected that the standard deviations of Final Average Academic Grades are somewhat lower than most of the standard deviations of the department grades. If it is desired that the standard deviation of Final Average Academic Grades be higher, it would be possible to achieve this by use of a different formula in arriving at the final averages.

Intercorrelations Among Bureau Achievement Examinations

Intercorrelations among Bureau examinations are presented in one part of Table 3.5, for Classes VI through IX. In Class VI, the two parts of the Navigation examination were entered as separate variables, while in the remaining classes they were combined into one variable according to the weighting formula in use at Newport.

In general, the correlations among the examinations are only of moderate size. The correlation between the Engineering and Naval Weapons examination is consistently the highest. It should be pointed out that the various revisions of the examinations did not seem to affect their intercorrelations to any significant extent, nor in any systematic way.

TABLE 3.4

RELATIVE ACTUAL WEIGHTS OF COMPONENTS OF THE
FINAL ACADEMIC AVERAGE
OCS CLASS VIII

Component	r with composite	β	Proportional weight
Engineering Dept. Grade	.852	.061	.052
Navigation Dept. Grade	.882	.332	.286
Operations Dept. Grade	.840	.211	.182
Orientation Dept. Grade	.759	.037	.032
Seamanship Dept. Grade	.781	.123	.106
Weapons Dept. Grade	.923	.260	.224
BuExam Average	.777	.138	.119
	.997 (multiple R)	--	1.000

TABLE 3.5

INTERCORRELATIONS AMONG FOURTEEN VARIABLES. OCS, NEWPORT, R.I. CLASSES VI, VII, VIII, IX
The N's are those listed in Table 3.2

VARIABLE	OCS CLASS	FINAL AVE	DEPT. FINAL GRADES E	N	OP	O	S	W	BuEXAM AVE	ACHIEVEMENT EXAMINATIONS (BuEX) E	N	OP	O	S	W
FINAL ACADEMIC AVERAGE	VII	1.00	.86	.84	.89	.71	.84	.91	.85	.64 A'*	.64 B	.66 B	.48 B	.50 B	.67 B
	VIII	1.00	.85	.88	.84	.76	.78	.92	.78	.58 A'	.56 B'	.62 B	.47 B	.50 B'	.64 C
	IX	1.00	.87	.86	.90	.75	.84	.92	.85	.65 A'	.64 B'	.64 C	.61 B	.56 B'	.68 C
ENG. DEPT. GRADE	VII	.86	1.00	.63	.72	.54	.75	.84	.75	.70 A'	.49 B	.55 B	.35 B	.44 B	.64 B
	VIII	.85	1.00	.74	.57	.68	.87	.83	.55	.50 A'	.33 B'	.41 B	.32 B	.35 B'	.47 C
	IX	.87	1.00	.66	.76	.62	.75	.85	.74	.68 A'	.47 B'	.18 C	.49 B	.48 B'	.65 C
NAV. DEPT. GRADE	VII	.84	.63	1.00	.72	.49	.59	.64	.64	.40 A'	.61 B	.53 B	.30 B	.30 B	.42 B
	VIII	.88	.74	1.00	.63	.59	.69	.71	.57	.39 A'	.53 B'	.46 B	.31 B	.31 B'	.41 C
	IX	.86	.66	1.00	.74	.53	.61	.67	.66	.44 A'	.64 B'	.53 C	.42 B	.39 B'	.46 C
OPERATIONS DEPT. GRADE	VII	.89	.72	.72	1.00	.63	.73	.79	.73	.52 A'	.58 B	.59 B	.44 B	.44 B	.54 B
	VIII	.84	.57	.63	1.00	.67	.48	.76	.75	.55 A'	.52 B'	.63 B	.51 B	.51 B'	.60 C
	IX	.90	.76	.74	1.00	.69	.75	.82	.75	.56 A'	.57 B'	.57 C	.55 B	.52 B'	.56 C
ORIEN. & MIL. JUSTICE DEPT. GRADE	VII	.71	.54	.49	.63	1.00	.71	.62	.55	.34 A'	.36 B	.45 B	.57 B	.39 B	.42 B
	VIII	.76	.63	.59	.67	1.00	.73	.69	.51	.31 A'	.35 B'	.43 B	.50 B	.38 B'	.39 C
	IX	.75	.62	.53	.69	1.00	.71	.68	.59	.40 A'	.38 B'	.46 C	.64 B	.45 B'	.44 C
SEAMANSHIP DEPT. GRADE	VII	.84	.75	.59	.73	.71	1.00	.76	.68	.50 A'	.45 B	.53 B	.49 B	.46 B	.56 B
	VIII	.78	.87	.69	.48	.73	1.00	.74	.41	.29 A'	.26 B'	.34 B	.29 B	.30 B'	.34 C
	IX	.84	.75	.61	.75	.71	1.00	.78	.68	.54 A'	.44 B'	.52 C	.56 B	.51 B'	.56 C
NAVAL WEAP. DEPT. GRADE	VII	.91	.84	.64	.79	.62	.76	1.00	.78	.66 A'	.52 B	.57 B	.44 B	.49 B	.67 B
	VIII	.92	.83	.71	.76	.69	.74	1.00	.72	.60 A'	.45 B'	.57 B	.41 B	.48 B'	.64 C
	IX	.92	.85	.67	.82	.68	.78	1.00	.79	.66 A'	.50 B'	.58 C	.55 B	.51 B'	.70 C
BUEXAM AVERAGE	VII	.85	.75	.64	.73	.55	.68	.78	1.00	.75 A'	.76 B	.73 B	.52 B	.56 B	.82 B
	VIII	.78	.55	.57	.75	.51	.41	.72	1.00	.75 A'	.71 B'	.75 B	.60 B	.63 B'	.85 C
	IX	.85	.74	.66	.75	.59	.68	.79	1.00	.75 A'	.73 B'	.69 C	.62 B	.61 B'	.82 C

*Successive major revisions of the original BuPers Officer Candidate Achievement Examination (Form A) are denoted by the letters B and C. Minor revisions are denoted by the prime.

Key: E = Engineering; N = Navigation; OP = Operations; O = Orientation; S = Seamanship; W = Naval Weapons

TABLE 3.5 (continued)

VARIABLE	OCS CLASS	FINAL AVE	DEPT. FINAL GRADES E	N	OP	O	S	W	BuEXAM AVE	ACHIEVEMENT EXAMINATIONS (BuEX) E	N**	OP	O	S	W
ENGINEERING EXAM	VI	*	*	*	*	*	*	*	*	1.00 A	.04 / .18 A	.49 A	.35 A	.44 A	.66 A
	VII	.64	.70	.40	.52	.34	.50	.66	.75	1.00 A'	.39 B	.49 B	.28 B	.43 B	.65 B
	VIII	.58	.50	.39	.55	.31	.29	.60	.75	1.00 A'	.34 B'	.54 B	.36 B	.44 B'	.65 C
	IX	.65	.68	.44	.56	.40	.54	.66	.75	1.00 A'	.40 B'	.43 C	.44 B	.47 B'	.65 C
NAVIGATION EXAM	VI	*	*	*	*	*	*	*	*	.04 / .18 A	1.00 / 1.00 A	.46 / .29 A	.32 / .14 A	.40 / .20 A	.48 / .24 A
	VII	.64	.49	.61	.58	.36	.45	.52	.76	.39 A'	1.00 B	.45 B	.28 B	.32 B	.40 B
	VIII	.56	.33	.53	.52	.35	.26	.45	.71	.34 A'	1.00 B'	.42 B	.32 B	.30 B'	.38 C
	IX	.64	.47	.64	.57	.38	.44	.50	.73	.40 A'	1.00 B'	.43 C	.33 B	.36 B'	.39 C
OPERATIONS EXAM	VI	*	*	*	*	*	*	*	*	.49 A	.46 / .29 A	1.00 A	.40 A	.43 A	.54 A
	VII	.66	.55	.53	.59	.45	.53	.57	.73	.49 A'	.45 B	1.00 B	.34 B	.38 B	.53 B
	VIII	.62	.41	.46	.63	.43	.34	.57	.75	.54 A'	.42 B'	1.00 B	.46 B	.49 B'	.56 C
	IX	.64	.18	.53	.57	.46	.52	.58	.69	.43 A'	.43 B'	1.00 C	.39 B	.37 B'	.50 C
ORIENTATION & MIL. JUST. EXAM	VI	*	*	*	*	*	*	*	*	.35 A	.32 / .14 A	.40 A	1.00 A	.36 A	.43 A
	VII	.48	.35	.30	.44	.57	.49	.44	.52	.28 A'	.28 B	.34 B	1.00 B	.25 B	.35 B
	VIII	.47	.32	.31	.51	.50	.29	.41	.60	.36 A'	.32 B'	.46 B	1.00 B	.39 B'	.41 C
	IX	.61	.49	.42	.55	.64	.56	.55	.62	.44 A'	.33 B'	.39 C	1.00 B	.44 B'	.47 C
SEAMANSHIP EXAM	VI	*	*	*	*	*	*	*	*	.44 A	.40 / .20 A	.43 A	.36 A	1.00 A	.52 A
	VII	.50	.44	.30	.44	.39	.46	.49	.56	.43 A'	.32 B	.38 B	.25 B	1.00 B	.43 B
	VIII	.50	.35	.31	.51	.38	.30	.48	.63	.44 A'	.30 B'	.49 B	.39 B	1.00 B'	.50 C
	IX	.56	.48	.39	.52	.45	.51	.51	.61	.47 A'	.36 B'	.37 C	.44 B	1.00 B'	.45 C
NAVAL WEAPONS EXAM	VI	*	*	*	*	*	*	*	*	.66 A	.48 / .24 A	.54 A	.43 A	.52 A	1.00 A
	VII	.67	.64	.42	.54	.42	.57	.67	.82	.65 A'	.40 B	.53 B	.35 B	.43 B	1.00 B
	VIII	.64	.47	.41	.60	.39	.34	.64	.85	.65 A'	.38 B'	.56 B	.41 B	.50 B'	1.00 C
	IX	.68	.65	.46	.56	.44	.56	.70	.82	.65 A'	.39 B'	.50 C	.47 B	.45 B'	1.00 C

* Not computed.

** For Class VI the upper correlation is for part 1; the lower figure is for part 2. The correlation between part 1 and part 2 was .27.

Since the correlations are somewhat obscured by the rather low reliabilities of the variables, the data for Class VIII were corrected for attenuation. That is, use was made of the standard formula for estimating the correlation between two perfectly reliable tests:

$$r_{\infty\infty} = \frac{r_{12}}{\sqrt{r_{11}}\ \sqrt{r_{22}}}$$

The results are shown in Table 3.6. It can now be seen that the various subject-matters covered by the tests are fairly homogeneous. Certainly, though no factor analysis has been made of these data, there is a general factor of achievement running through the tests. Engineering and Naval Weapons would have the highest loading on this general factor, and Navigation and Orientation would have considerably smaller loadings. There is some evidence for a group factor which would include Seamanship and Operations. In any case, Navigation is the most independent of the examinations. Its relatively high correlation with Operations (particularly evident in the fact that Part II of Navigation correlated most highly with Operations in Class VI) is probably due to the fact that both examinations have the common feature of plotting--navigational plotting and maneuvering board plotting.

TABLE 3.6

INTERCORRELATIONS OF BUREAU EXAMINATIONS, CORRECTED FOR ATTENUATION, OCS CLASS VII(GRADUATING DECEMBER 1952)

The reliabilities used in these computations are given in the last column of this table.

N = 772

	Eng.	Nav.	Oper.	Orient.	Seam.	Weap.	rel.
Engineering	-	.464	.774	.559	.698	.874	.743
Navigation	.464	-	.610	.509	.481	.518	.715
Operations	.774	.610	-	.766	.816	.794	.669
Orientation	.559	.509	.766	-	.720	.645	.546
Seamanship	.698	.481	.816	.720	-	.796	.541
Naval Weapons	.874	.518	.794	.645	.796	-	.746

The correlations between the separate examinations and the "Bureau examination average" are also presented in Table 3.5. The magnitude of these correlations (which are part-whole correlations) is a function of the weights used in arriving at the composite average score, as well as the intercorrelations of the variables, and (indirectly) their reliabilities. The Naval Weapons examination consistently shows the highest correlation with the Bureau Examination average but it also has the highest nominal weight in the composite.

Intercorrelations Between Academic Department Grades

Table 3.5 also presents intercorrelations among academic department grades (including the final academic average) for Classes VII, VIII, and IX. These correlations tend to be higher, on the average, than the correlations among the Bureau examinations. The reliabilities of the department grades are not known, but may be assumed to be quite high in as much as they are based on a large number of quizzes, examinations, and performance ratings. Again, there is evidence of a general factor of achievement, and again, the correlation between Engineering and Naval Weapons is consistently the highest, no doubt because both these subjects involve engineering and mechanical training. In almost all respects, the pattern of correlations parallels in a general way that of the correlations among Bureau examinations. It may be concluded that achievement at OCS, as measured by academic grades, is a fairly homogeneous entity.

Intercorrelations Between Academic Grades and Bureau Examinations

Table 3.5 also presents intercorrelations between academic department grades and the several Bureau examinations, for OCS Classes VII, VIII, and IX.

If there is anything specific in the content of instruction in the several departments, it might be expected that the correlations between the academic grades for a given department and the corresponding Bureau examination would be

higher than correlations between non-corresponding departments and examinations. Such an expectation is borne out by the present data to only a limited extent. For example, the Engineering department grade is almost as highly correlated with the Naval Weapons examination as with the Engineering examination. The Operations department grade is correlated with all examinations almost equally; this is also true of the Seamanship department grade. Only in the case of the Navigation and Orientation and Military Justice department grades do we find fairly clear evidence of specificity. For example, in Class IX the correlation between the Navigation department grade and the Bureau examination is .64, which surpasses the next highest correlation .53, with the Operations examination. Likewise, in the same class the Orientation and Military Justice grade correlates .64 with the corresponding examination, clearly surpassing the correlation of .56 with the Operations examinations.

These data, then, support the impression gained earlier that if the curriculum is considered by departments, the material can be classified in three types:

I Engineering and Naval Weapons material -- involving mechanical and engineering training.

II Navigation -- involving ability in mathematics (particularly Trigonometry) and graphical plotting.

III Orientation and Military Justice -- involving ability to acquire factual information about navy life, customs, and regulations.

In the meantime, subject-matter considered under Seamanship and Operations has a somewhat ambiguous status. From the standpoint of factor analysis, it could be said that these subjects have a complex factor composition.

The process of revising the tests seemed to have no systematic effect on their correlations with department grades. If only the data for Classes VII and VIII had been considered, it might have been said that the correlations in Class VIII

were in some cases quite different from those of Class VII, and some credence would have been lent to the belief that the revisions of the examinations were responsible for these changes. However, the data for Class IX belie such an interpretation, for in most respects the pattern of correlations was quite similar to that for Class VII. There was, however, one surprising change to be noted in the correlation of the Engineering department grades and the Operations examination. In Classes VII and VIII, where Form B of the Operations examination was used, these correlations were .55 and .41, respectively. In Class IX, where a radically different form of the examination was used (Form C) the correlation dropped to .18. No simple explanation for this effect seems apparent, however, computations have been checked carefully. Furthermore, Form C maintained its previous correlations with other department grades.

Item Analyses of OCS Examinations

In the process of constructing revised forms of OCS examinations, item analyses of many of the previously prepared examinations (both for NROTC and OCS) were made available by the Bureau of Naval Personnel. Much attention was paid to these item analysis data wherever there was a question of selecting an item from a previous examination for use in one of the new examinations. That is, an attempt was made to avoid very easy items and items of low validity (validity in this case being defined as the correlation of the item with total score).

Since it was eventually going to be necessary to select items from the OCS examinations for use in the May 1953 examinations, item analyses were made of these examinations as they were developed. All examinations administered to Classes VII and VIII were subjected to this analysis, as well as Form C of the Operations examination administered to Class IX, since this was a new form.

Three types of data were secured on each item:

(1) The number and the proportion passing the item. The proportion passing was identified as the Difficulty Index.

(2) The estimated correlation between the item and the Bureau examination average (a composite score described previously on page 38). This correlation was determined on the basis of Flanagan's procedure involving groups of 9, 20, 42, 20, and 9 per cent.* Facilitating tables for these determinations have been prepared by Flanagan. As noted in Chapter II, the composite score on the examinations was selected as the best available criterion of "integrated knowledge."

(3) Frequencies of responses to the five alternatives in each item.

The data resulting from these item analyses have been entered on the item cards supplied separately to the Bureau of Naval Personnel, and since they are scarcely meaningful except when viewed in conjunction with the tests themselves, they are not included in detail in this report. The results are, however, summarized in Table 3.7, where medians and the 10 and the 90 percentile points of the distributions of difficulty indices (per cent correct) and item validities are shown.

---

* The item analyses were based on a random 33% sample of the classes. Thus, the numbers of cases included in each of the groups were as follows:

	Class VII	VIII	IX
Top 9%	21	24	25
Next 20%	46	52	56
(Middle 42%)	(94)	(105)	(116)
Lower 20%	46	52	56
Bottom 9%	21	24	25
Sample Total	228	257	278
Class Total	684	772	834

The parentheses in the above table indicate that the middle group is not actually used in Flanagan's procedure.

TABLE 3.7

SUMMARY OF ITEM ANALYSIS RESULTS, OCS EXAMINATIONS

(CLASSES VII, VIII, AND IX)

Examination	Class	Form	Difficulty Index P_{10}	Difficulty Index Mdn	Difficulty Index P_{90}	Item Validity P_{10}	Item Validity Mdn	Item Validity P_{90}	No. of Items
Engineering	VII	A'	.41	.59	.87	.08	.30	.51	75
Engineering	VIII	A'	.41	.66	.91	.06	.25	.38	75
Navigation Part I	VII	B	.42	.81	.95	.10	.22	.38	30
Navigation Part I	VIII	B'	.38	.80	.96	.06	.26	.38	30
Navigation Part II	VII	B	.37	.87	.99	.16	.36	.50	30
Navigation Part II	VIII	B'	.38	.88	.99	.00	.24	.41	30
Operations	VII	B	.34	.69	.93	.03	.26	.38	75
Operations	VIII	B	.32	.71	.93	.02	.20	.33	75
Operations	IX	C	.44	.62	.86	.06	.19	.35	50
Orientation	VII	B	.39	.74	.94	-.03	.14	.30	75
Orientation	VIII	B	.36	.75	.95	-.04	.14	.26	75
Seamanship	VII	B	.54	.86	.97	-.03	.15	.37	50
Seamanship	VIII	B'	.41	.68	.94	.03	.19	.33	50
Naval Weapons	VII	B	.39	.76	.93	.06	.27	.44	75
Naval Weapons	VIII	C	.34	.67	.90	.09	.27	.45	75

Several points must be held in mind in considering the item validity data. First, the median item validities may be partly a function of the weight of the examination in the composite score which was used as the criterion. For example, the rather high validities found in the Naval Weapons Examination may result from the relatively high weight (a nominal weight of 2) of that examination in the composite. Secondly, it cannot necessarily be expected that the median item validity of a revised test will be appreciably higher then the median item validity of a previous test, even when item analysis data were used in the revision, in view of the fact that there is an inevitable regression effect due to sampling errors in the item validity statistics. Furthermore, most of the revisions which took place in the OCS examinations represented attempts to try out numerous new items never before used.

It should be noted that since the tests were either not at all speeded or only slightly speeded, any tendency for items near the end of the test to have spuriously raised validities due to time limits was negligible.

With these points to be borne in mind, Table 3.7 suggests the following conclusions:

(a) The range of items in difficulty is in general satisfactory. For this type of examination it is felt desirable to include items of a wide range of difficulty, in a distribution which is more nearly like a rectangular distribution than a normal distribution. The distributions of item difficulties in a number of the present examinations are fairly rectangular (in Engineering, Operations, Orientation, and Naval Weapons). Especially in the Navigation examination, however, the distribution of item difficulties tends to be negatively skewed, with a high median. Apparently this subject is so well and thoroughly taught that even the relatively difficult problems set in the examination are solved by a large number of students. It should be noted that despite these characteristics of the item-

difficulty distribution, the raw score distributions are approximately normal, and the tests are among the more reliable in the battery. Indeed, it will be recalled that Navigation, Part II had consistently the highest per-item reliability. These results only serve to emphasize again (as has been shown in recent literature on the theory of mental tests) that the form of distribution of item-difficulties is only a small factor in the total reliability of a test.

(b) The median item validities with the chosen criterion are about as satisfactory as may be expected for tests of this type. There is a tendency for examinations emphasizing engineering and mathematical problems (such as Engineering, Navigation, and Naval Weapons) to have higher median item validities than examinations emphasizing factual information (such as Orientation and Seamanship). This probably reflects the fact that the criterion score is most heavily weighted with variance from examinations of the former type. The fact that an item has a low validity against the criterion should not necessarily discredit the item, for it may be measuring another type of achievement than whatever is represented by the composite criterion. It was with this hope that the further statistical analyses described in Chapter IV were undertaken.

As a final point worthy of mention, it may be reported that the project staff saw no indication that item validities were a function of item type. While some of the low item validities were clearly due to ambiguities in statement, high validities were found equally often for items appearing to measure sheer factual information and for items appearing to require problem-solving and decision making.

CHAPTER IV

FACTOR ANALYSIS OF THE OCS ACHIEVEMENT EXAMINATIONS

Introduction

One of the fundamental questions for which some kind of answer was desired in the present research was whether the knowledge shown by officer candidate examinees at the end of their training is of a homogeneous or a heterogeneous character. If the former, it would be pointless to hope that "integrated" examinations (or any other type of examination, for that matter) measuring something quite different from the previous examinations could be constructed, unless one makes the asumption that previous examinations failed utterly to measure "integrated" knowledge. If the latter, there arises the problem of identifying the separate aspects of achievement tested by the examinations. It might also be hoped that one or more of these separate aspects of achievement might be of such a nature as to be regarded as measuring "integrated knowledge" rather than merely factual information.

This chapter reports an attempt to find and identify the separate aspects of achievement measured by the OCS achievement examinations. All data reported on here are based on the examinations given to OCS Class VIII in December 1952. These examinations, with the exception of the Operations examination, (which was subsequently revised for the March 1953 class), represented the latest efforts of the project as far as OCS examinations were concerned.

Selection of a Representative Set of 72 Items for Factor Analysis

The method selected as most practicable and useful for identifying separate aspects of achievement was that known as the Wherry-Gaylor iterative factor

analysis procedure.* This is a variety of factor analysis procedure which is especially designed for use with tests composed of large numbers of items. In theory, it is a workable method, but in practice it was found to present certain difficulties, chiefly difficulties which stem from the fact that the criteria for determining iterations are not completely objective and are affected unduly by sampling errors in item-test correlations.

In the OCS Class VIII examinations, 410 items in all were available for analysis. In view of the fact that the Wherry-Gaylor procedure is quite laborious, it was decided to analyze only a representative set of items. Seventy-two items were therefore drawn in such a way as to fit into all of the 72 possible combinations of the following four dimensions:

1. Subject-Matter. (Four Categories)
   - A. Operations
   - B. Navigation
   - C. Naval Weapons
   - D. Orientation and Military Justice

     (Three items from Seamanship had to be used because the Navigation examination did not yield all of the 18 items required by this design.)

2. <u>Item-Type</u>. (Two categories)
   - A. Items judged by project personnel to depend chiefly on memory of factual information, terminology, special naval rules of procedure, etc.

---

* Wherry, R. J., and Gaylord, R. H. The concept of test and item reliability in relation to factor pattern. Psychometrika, 1943, 8, 247-269. Unfortunately, a revision of this procedure was published too late for use in this project (Wherry, R. J., and Winer, B. J. A method for factoring large numbers of items. Psychometrika, 1953, 18, 161-179).

B. Items judged by project personnel to depend chiefly on higher mental processes such as reasoning, problem solving, decision-making, comparison, evaluation, computation, etc., with the memory element at a minimum.

3. Difficulty. (Three categories)

After the items in the given subject-matter and of a given item-type were identified, items were selected in three ranges of item difficulty:

A. Low Difficulty

B. Medium Difficulty

C. High Difficulty

4. Item Validity. (Three categories)

The item-analysis data described in Chapter III were used in selecting items in three ranges of item validity:

A. Low Validity (including near-zero validity)

B. Medium Validity

C. High Validity

The items were drawn chiefly from the four examinations which had been most worked on by the project staff and hence were believed most likely to contain a variety of items differentiable in terms of various aspects of achievement. The ranges of item difficulty and validity values were set up after items had been selected for subject-matter and item-type; these ranges differed slightly among the four subject-matter categories. The items were classified by item-type at a conference of the three professionals on the project staff who had worked in the construction of the examinations and who were hence fairly familiar with the content and nature of each item.

Table 4.1 lists the 72 items selected for the iterative factor analysis. Items are identified in terms of (1) item number in the set of 72 items -- this set was known as "Key TOOO," and (2) item number by examination of the Class VII OCS examinations.

TABLE 4.1

LIST OF 72 REPRESENTATIVE ITEMS IN THE OCS CLASS VIII EXAMINATIONS SUBJECTED TO ITERATIVE FACTOR ANALYSIS

Item No., Key TOOO	Item No. Dec. '52 OCS Exam	Item Type	Validity Classification	r	Difficulty Classification	D
			ORIENTATION AND MILITARY JUSTICE			
1	9	B	High	.25	Low	.36
2	16	A	High	.37	High	.85
3	17	A	Low	.06	High	.91
4	20	A	Low	.07	Med.	.81
5	21	A	Med.	.08	Low	.55
6	24	A	Low	.08	Low	.55
7	26	A	Med.	.09	Med.	.78
8	31	A	High	.25	Low	.35
9	44	B	Med.	.10	High	.99
10	51	B	Low	.00	Low	.42
11	61	A	High	.24	Med.	.72
12	63	B	Med.	.17	Low	.45
13	65	A	Med.	.16	High	.89
14	69	B	Low	-.05	High	.92
15	70	B	Med.	.18	Med.	.82
16	71	B	Low	.01	Med.	.64
17	72	B	High	.22	Med.	.70
18	75	B	High	.21	High	.97
			OPERATIONS			
19	1	A	Med.	.18	Med.	.68
20	2	A	Low	.11	High	.93
21	4	B	Low	.07	Low	.38
22	11	A	High	.41	Med.	.72
23	14	A	Low	.08	Med.	.74
24	18	B	High	.26	High	.84
25	25	B	Med.	.22	Med.	.72
26	28	B	Low	.04	Med.	.73
27	29	A	Med.	.18	High	.90
28	30	B	High	.40	Med.	.52
29	31	B	High	.31	Low	.45
30	36	A	Low	.02	Low	.45
31	40	A	High	.36	Low	.32
32	44	A	High	.31	High	.86
33	60	A	Med.	.20	Low	.27
34	70	B	Med.	.18	Low	.30
35	71	B	Med.	.15	High	.91
36	75	B	Low	.14	High	.80

Iterative Factor Analysis of 72 Items

The Wherry-Gaylord procedure was then applied to these 72 items. The data studied were the responses of a random sample of 200 cases from Class VIII. This size of sample was selected as sufficient in view of the fact that the procedure had previously been used successfully on 231 cases by Wherry, Campbell, and Perloff.*

The item-test correlations used in this procedure were exclusively tetrachoric correlations using a 50-50 criterion dichotomy. A special table of tetrachoric correlations based on the 50-50 dichotomy was prepared, in as much as the chart published by Adkins** could not be read with sufficient precision. Correlations which could not be read from the Thurstone tables*** were computed by the cosine formula.

The procedure resulted in five factors. The final keys which resulted were identified as S105, S207, S309, S401, and S501. The item-correlations with these five keys are shown in Table 4.2, which also specifies the items included in the keys. In obtaining scores on the keys, item scores were added with unit weights. The characteristics of the scores their intercorrelations, and the possible meaning of the factors will be discussed at a later point in this report, in conjunction with other results.

---

* Wherry, R. J., Perloff, R., and Campbell, J. T. An empirical verification of the Wherry-Gaylord iterative factor analysis procedure. Psychometrika, 1951, 16, 67-74.

** Page 184 in Adkins, D. C., The Construction and Analysis of Achievement Tests. Washington. D.C., Government Printing Office, 1947.

*** Chesire, L., Saffir, M., and Thurstone, L.L. Computing Diagrams for the Tetrachoric Correlation Coefficient. Chicago: Univ. Chicago Bookstore, 1933.

TABLE 4.2

ITEM-KEY CORRELATIONS FOR 72 ITEMS

SELECTED FOR ITERATIVE FACTOR ANALYSIS

All entries have been multiplied by 100 to eliminate the decimal point.

No. on Dec. OCS Exam**		Item # Key T000	Factor Key S105	S207	S309	S401	S501
O	9	1.	-17	45*	13	07	03
O	16	2.	08	52*	49*	-08	-02
O	17	3.	54*	-05	34*	05	13
O	20	4.	40*	-12	03	12	22
O	21	5.	38*	00	00	16	07
O	24	6.	30*	00	-02	-08	05
O	26	7.	12	12	09	00	34*
O	31	8.	23	13	07	46*	-03
O	44	9.	61*	100	62	25	-25
O	51	10.	42*	08	05	-08	26
O	61	11.	14	05	39	07	03
O	63	12.	17	-05	05	42*	07
O	65	13.	30	11	22*	00	11
O	69	14.	-05	13	-05	25*	100*
O	70	15.	35	05	55*	15	30
O	71	16.	-08	-12	02	08	42*
O	72	17.	50*	13	50*	13	03
O	75	18.	00	00	42	-42	-42
OP	1	19.	02	22	15	45*	-18
OP	2	20.	13	-05	05	34*	-13
OP	4	21.	-03	02	07	43*	38
OP	11	22.	16	42*	56*	-08	16
OP	14	23.	05	12	16	-02	-05
OP	18	24.	42*	32	52*	02	15
OP	25	25.	18	52*	47*	18	-10
OP	28	26.	46*	20	25*	12	00
OP	29	27.	13	05	05	13*	34
OP	30	28.	37	40	66*	28	10
OP	31	29.	36	32*	66*	03	-12
OP	36	30.	03	-11	-09	-14	51*
OP	40	31.	24	51*	31*	02	20
OP	44	32.	14	05	27	-08	50*
OP	60	33.	02	15	00	59*	-12
OP	70	34.	47*	-05	02	16	08
OP	71	35.	36*	09	00	-08	00
OP	75	36.	24*	-02	10	-19	00

* Denotes items which formed the corresponding "S" key.

** Key: O = Orientation and Military Justice
OP = Operations

TABLE 4.2 (continued)

No. on Dec. OCS Exam**		Item # Key T000	Factor Key S105	S207	S309	S401	S501
S	30	37.	12	63*	51*	20	-18
S	41	38.	54*	00	27*	18	08
S	44	39.	25	00	00	00	40*
N	1	40.	00	66*	40*	00	45
N	4	41.	18	18	52*	02	-02
N	5	42.	06	40	29	06	17
N	7	43.	03	18	37	18	22
N	16	44.	13	53*	43	17	-10
N	18	45.	02	14	18	60*	-24
N	23	46.	00	15	13	13	62*
N	26	47.	25	-34	12	-18	34*
N	27	48.	07	07	03	-18	50*
N	37	49.	-05	25	34*	-05	13
N	45	50.	-52	36	17	34	34
N	46	51.	-03	00	-07	42*	-16
N	52	52.	12	12	52*	08	08
N	58	53.	50*	14	56*	-05	05
N	60	54.	06	08	44*	-04	-11
W	3	55.	23	39*	36*	07	20
W	4	56.	02	-10	03	02	28*
W	5	57.	42	-42	00	100	-42
W	11	58.	35*	03	12	12	16
W	15	59.	03	07	07	18*	00
W	16	60.	10	07	07	52*	-03
W	17	61.	67*	09	16	09	24
W	27	62.	38*	26	45*	-38	-04
W	31	63.	14	14	35*	00	05
W	32	64.	-07	-10	16	31*	02
W	38	65.	-13	47*	16	07	07
W	39	66.	00	23	26	30*	-07
W	49	67.	21	64*	37	18	-06
W	53	68.	16	16	16	-07	56*
W	55	69.	-06	02	13	46*	12
W	57	70.	28	41	37	12	-10
W	68	71.	05	-05	-39	-10	61
W	69	72.	02	50*	14	02	-14

* Denotes items which formed the corresponding "S" key.

** Key: S = Seamanship
N = Navigation
W = Naval Weapons

Expansion of the Factor Keys

Since the S-keys contained relatively few items, the scores were then used as criteria for identifying other items among the 410 items available which could be considered as measuring the same factors. The same sample of 200 cases was used as for the previous analysis. The item-key correlations, again computed by the special tetrachoric correlation tables involving a 50-50 dichotomy, are shown in those columns of Table 4.3 which are headed by the number of one of the S-keys. In the case of items which appeared on S-keys, the item-key correlations may not be the same as those shown in Table 4.2 because of the fact that the upper and lower 50 per cent groups on the criterion groups were sorted in a slightly different way.*

Items were then selected for the expanded keys by the following rules:

(a) All items on the original S-keys were retained, except as provided by rule (c) below.

(b) New items must have a correlation of .30 or greater with the S-key.

(c) The item could be assigned to no more than one E-key, that is, to the factor where its item-key correlation was highest.

(d) Very easy items showing unstable item-key correlations were to be ignored.

The results of this "Operation Dragnet" were disappointing. Relatively few new items were added to the S-keys to form the E-keys. In fact, all five E-keys included only 136, or 33 per cent of the 410 items. Perhaps the requirement that the correlations had to be .30 or greater was too rigorous. Correlations of .30 or greater were relatively infrequent in the data. But it must be remembered that an attempt was being made to identify separate aspects of achievement; hence it was necessary to be assured of clarity of the results in order to make interpretations. It had been thought that many new items in the total of 410 available would fall into one of the E-keys.

---

* In every case, the papers having a score at the median were randomly assigned to the upper and lower groups; this random assignment was usually different each time the score was used as a criterion.

TABLE 4.3

CORRELATIONS OF ALL ITEMS

IN DECEMBER OCS EXAMINATIONS WITH "S" KEYS AND "E" KEYS

Note: * Denotes that item was selected for the corresponding E Key.
** Denotes that item was in the original "S" Key.
*** Denotes item on S Key and on E Key.

All entries have been multiplied by 100 to eliminate decimal point.

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
ORIENTATION EXAMINATION									
1.	05	22	05	-05	02	02	05	24	12
2.	10	10	-10	-10	-10	-10	-10	10	-10
3.	17	-12	-12	-08	-05	-16	02	00	12
4.	10	00	27	27	31*	39	00	-05	-05
5.	20	10	13	17	17	13	00	08	-07
6.	20	30	-05	09	16	12	30*	59	12
7.	00	100	100	-100	-100	-100	00	00	00
8.	20	04	12	35	04	20	60*	73	-04
9.	-13	-10	45***	33	03	10	13	17	-03
10.	-27	-21	-20	-21	-27	-27	-21	-21	21
11.	22	-13	02	10	18	06	13	11	06
12.	16	08	08	26	08	20	08	16	-15
13.	08	08	02	-08	18	13	18	18	02
14.	00	15	00	-05	10	05	-10	15	-05
15.	04	04	16	10	30*	16	-24	-10	04
16.	08	22	44**	33	48***	48	02	08	02
17.	40***	38	05	00	34**	12	05	-13	25
18.	10	17	13	03	07	-07	-07	02	00
19.	10	10	23	17	23	10	03	13	07
20.	30***	40	-15	-07	-08	02	08	-08	25
21.	33***	38	03	00	03	03	07	16	07
22.	15	28	21	12	24	24	-05	05	12
23.	-09	17	00	-09	-05	-05	00	00	30*
24.	26**	05	00	-05	05	05	-12	-14	00
25.	-10	-10	15	18	03	13	13	03	15
26.	12	12	12	00	05	-05	00	-12	34***
27.	100	-100	-100	-100	100	100	100	100	100
28.	10	04	16	22	53*	48	10	-16	16
29.	02	05	05	02	-08	-05	-02	08	-08
30.	13	07	03	10	24	22	07	12	-14
31.	20	27	17	37	-03	03	30***	38	-03
32.	-02	-02	22	22	18	04	-02	-08	29
33.	-24	-32	12	08	-08	05	-02	02	-20
34.	-09	-02	-14	-02	05	02	09	20	05
35.	13	22	00	07	00	07	07	08	16

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
36.	-18	-08	12	18	05	05	-05	12	-28
37.	-10	-07	18	12	-03	10	10	07	-07
38.	16	12	04	00	20	08	12	08	-12
39.	18	08	-18	-18	-18	-18	18	18	02
40.	32*	42	08	03	22	18	05	-03	15
41.	09	12	16	05	20	20	12	-02	02
42.	20	20	04	16	08	04	00	08	-04
43.	10	24	24	16	16	-02	24	02	24
44.	42**	25	100	42	24	42	42	25	-25
45.	06	22	11	06	06	06	06	-14	11
46.	21	21	34*	43	18	24	18	16	-02
47.	14	-02	02	-02	14	02	06	-02	14
48.	02	14	12	08	18	12	12	05	02
49.	-26	-18	00	05	-18	-10	22	18	-18
50.	08	08	26	38	08	08	05	26	-15
51.	42***	56	14	02	08	-11	-11	-11	20
52.	10	-03	00	18	00	-07	-07	-07	07
53.	22	39	39	46	15	09	54*	86	15
54.	25	04	18	25	18	25	18	04	12
55.	-05	05	-05	-09	-05	13	25	44	-05
56.	21	66	00	21	42*	42	00	00	00
57.	-04	-04	04	12	-12	-16	-04	16	12
58.	10	10	38*	40	28	32	07	07	12
59.	03	16	-10	-16	00	-10	07	03	-13
60.	-02	08	-02	-05	-05	03	05	-05	-02
61.	17	07	10	24	28	30	07	17	07
62.	-30	00	00	-09	-22	-22	00	30	-15
63.	08	11	11	-05	08	08	48***	44	08
64.	-03	00	03	17	17	17	24	15	00
65.	30*	57	11	-06	22**	11	05	-05	11
66.	10	13	18	10	10	10	18	12	13
67.	-100	-100	-100	-100	100	100	100	100	-100
68.	-34	07	07	07	-07	20	19	07	-34
69.	-13	-05	13	-13	-05	-05	34**	-12	69***
70.	30	24	10	10	60***	50	20	00	45
71.	-05	08	-15	-02	-02	-08	15	10	32***
72.	47**	27	20	27	53***	43	03	13	10
73.	00	13	-08	08	-08	-08	20	20	20
74.	42	00	00	-42	-42	-42	00	--	-42
75.	42	42	00	00	-42	00	00	00	-42

OPERATIONS EXAMINATION

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
1.	-02	07	15	08	05	07	42***	58	-25
2.	05	36	-05	-38	05	-12	34***	53	-13
3.	10	10	06	-06	06	02	18	08	-14
4.	00	12	03	00	03	03	40**	25	40***
5.	05	14	05	02	14	12	28	19	12

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
6.	03	16	25	28	34*	38	16	24	07
7.	-09	-02	18	20	18	12	02	12	-24
8.	07	15	28	28	48*	42	18	22	11
9.	16	03	03	07	10	07	07	07	-03
10.	03	03	23	23	07	13	03	17	00
11.	17	18	39**	30	52***	59	-05	18	08
12.	07	07	24	24	42*	50	16	20	-02
13.	16	08	31*	44	21	44	21	36	05
14.	02	-05	05	12	16	12	02	-05	-12
15.	05	22	08	15	22	25	-02	05	12
16.	--	00	--	--	--	--	--	--	--
17.	18	08	12	24	30*	28	12	18	-03
18.	42**	36	32	28	42**	42	14	11	07
19.	-24	24	42	42	42	100	14	36	25
20.	23	12	33*	23	29	23	02	50	-18
21.	00	12	30	49	41*	64	30	00	00
22.	15	17	-08	05	-02	-08	20	18	12
23.	15	15	30	46	30	46	09	22	00
24.	09	16	-16	-09	-09	-09	24	30	00
25.	18	21	52***	47	49**	32	10	03	00
26.	24	20	24	40	40*	51	24	44	04
27.	03	07	19	13	22	22	-03	-02	00
28.	42***	34	05	09	39**	25	12	00	10
29.	34	13	05	-05	-05	-13	34**	-05	25**
30.	31	24	43	40	54***	60	31	40	13
31.	32	13	55	55	63***	71	-03	02	-12
32.	100	100	100	100	100	-100	100	-100	100
33.	09	12	21	24	17	28	21	22	-08
34.	03	08	18	18	18	29	-08	12	-12
35.	-05	05	12	18	12	12	05	12	05
36.	-03	11	-17	-14	-08	-17	03	-02	45**
37.	-05	05	02	-02	-02	-05	-08	-03	05
38.	-05	-05	20	10	10	05	05	05	-05
39.	18	24	22	10	22	18	-03	12	03
40.	09	16	48***	48	41**	28	05	05	28
41.	100	100	-100	--	100	--	100	100	100
42.	-16	00	00	-03	03	03	-03	-07	07
43.	15	18	08	08	-08	15	02	15	-12
44.	16	20	16	08	26	30	02	-02	48***
45.	-05	14	10	05	19	29	14	36	10
46.	27	27	08	08	14	18	08	03	18
47.	07	07	07	03	-03	00	18	15	07
48.	09	09	18	22	13	26	-13	18	05
49.	07	00	15	19	15	30	22	08	-03
50.	00	19	07	10	19	19	14	10	-10
51.	07	10	-03	03	10	07	07	07	-23
52.	14	-02	14	26	06	10	14	00	06

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
53.	02	-08	20	32	24	14	08	14	-05
54.	21	31	-02	12	24	18	34*	42	15
55.	-15	-12	-05	02	02	02	-02	10	02
56.	-12	-12	12	00	07	-12	20	26	00
57.	17	17	34	34	34	34	-17	17	34
58.	00	07	07	07	16	25	03	00	-10
59.	-18	-13	22	28	-03	07	10	26	-22
60.	-05	10	12	12	-08	-05	49***	52	-26
61.	00	22	05	08	-15	-05	05	02	-19
62.	05	10	16	27	10	10	05	16	-05
63.	02	-02	22	32	-02	02	22	40	-05
64.	20	16	17	30	05	02	11	10	-02
65.	05	34	-05	00	-30	-24	-24	00	-24
66.	16	20	18	-04	-06	-04	-33	-37	04
67.	45	30	10	04	10	04	45	24	30
68.	-13	13	13	25	-05	-05	05	05	13
69.	34*	34	13	08	08	-12	00	08	08
70.	47***	60	-05	16	05	02	08	08	19
71.	40***	69	18	18	00	00	-11	11	11
72.	27	36	18	11	00	-11	11	27	00
73.	23	23	26	23	20	30	10	23	03
74.	18	11	14	32	18	18	-02	14	-07
75.	33***	19	-14	05	10	15	-10	15	00

SEAMANSHIP EXAMINATION

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
1.	-42	25	25	42	-25	00	42	42	00
2.	00	00	00	-21	42	21	-66	00	-42
3.	03	18	16	10	00	13	22	28	18
4.	23	14	26	28	39*	35	08	08	-02
5.	03	-03	07	03	07	00	07	18	-03
6.	08	18	18	02	24	24	02	08	13
7.	10	-03	18	31	22	13	07	13	00
8.	09	02	-16	-12	02	-05	20	09	-05
9.	-40	-40	12	07	07	12	07	26	-12
10.	09	09	05	20	-09	11	18	34	-14
11.	-25	25	24	-62	-24	-24	-13	24	-25
12.	27	30	11	18	14	05	-05	11	05
13.	29	12	19	05	32*	38	05	13	02
14.	-06	-02	22	06	10	06	-13	-02	02
15.	12	15	09	18	15	05	-02	07	12
16.	-16	-05	-05	-12	02	-12	12	24	-12
17.	16	34	-04	-10	04	10	26	34	-04
18.	08	20	34*	32	20	34	00	00	-13
19.	22	18	28	18	18	22	10	14	-14
20.	16	10	16	27	10	00	16	16	00

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
21.	27	05	10	16	10	00	-10	-05	-10
22.	-16	-18	12	03	-03	00	10	12	07
23.	24	30	24	00	12	00	-12	00	00
24.	03	08	08	08	14	08	08	-03	08
25.	18	13	18	24	18	24	-08	13	-24
26.	34*	52	29	25	25	20	07	11	02
27.	13	02	24	28	18	18	18	-02	28
28.	05	-05	24	24	29	29	05	00	-10
29.	10	25	03	07	07	10	13	10	07
30.	18	14	62***	62	54**	45	12	20	-20
31.	12	18	11	12	-10	03	18	12	00
32.	13	24	10	08	28	24	13	-02	12
33.	100	100	00	00	00	00	00	00	100
34.	16	19	07	10	03	16	22	22	-19
35.	05	15	-10	-12	-15	08	00	08	-15
36.	13	10	13	13	13	00	-12	00	08
37.	-13	-07	10	-10	00	13	12	13	00
38.	25	05	25	25	-13	-05	25	38	-13
39.	-05	-02	-09	-20	-02	-14	-05	-05	08
40.	-07	10	33*	31	31	22	03	22	-10
41.	50**	55	18	27	36**	18	00	18	18
42.	16	22	18	10	28	31	00	-28	25
43.	09	02	19	09	05	05	12	05	02
44.	14	28	-14	-14	00	00	14	-14	43***
45.	-08	03	18	03	03	18	-02	-08	13
46.	09	28	02	09	09	12	02	02	02
47.	09	05	24	20	15	17	-02	05	02
48.	12	-22	02	02	09	05	-08	09	-02
49.	05	15	15	12	15	-02	05	08	-02
50.	10	10	16	16	10	10	04	16	-22

NAVIGATION EXAMINATION

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
1.	00	-11	50**	55	36**	27	11	-11	55
2.	05	-05	12	02	02	12	12	12	08
3.	07	12	30	40	53*	65	12	12	00
4.	28	14	11	22	45***	44	-02	02	02
5.	15	08	38*	40	22	20	06	-04	20
6.	-09	07	12	12	07	07	18	15	05
7.	00	10	20	24	31*	34	13	10	22
8.	11	22	-11	-11	-11	11	00	00	-22
9.	22	22	22	30	40	40	40	22	-11
10.	07	10	22	18	22	31	16	22	03
11.	18	24	15	20	05	15	15	12	05
12.	03	00	-03	-03	07	-03	-07	-19	13
13.	08	02	15	26	22	26	15	26	08
14.	27	27	-21	-21	100	27	27	21	21
15.	10	-03	36*	42	20	18	14	24	-14

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501
16.	02	13	57***	60	43	42	24	29	-07
17.	16	40	30*	30	00	-04	09	-04	00
18.	02	14	23	24	18	12	58***	51	-16
19.	-100	-100	100	100	-100	100	-100	-100	100
20.	05	17	42*	42	20	12	12	36	-08
21.	-05	-25	05	25	00	05	05	13	-05
22.	13	20	05	20	20	13	-26	-05	12
23.	24	36	24	-24	-24	-24	14	24	100**
24.	30*	40	09	09	09	24	-09	-09	-16
25.	10	-03	33*	37	14	26	14	18	-03
26.	25	12	-25	-12	05	18	-18	05	48***
27.	00	07	07	-03	14	03	-18	-30	43*
28.	29	18	14	03	29	32	10	14	-03
29.	18	18	28	34	31*	42	08	18	-05
30.	00	24	07	13	-03	03	10	03	22
31.	16	16	24	05	32	32	-24	-02	58*
32.	07	00	18	22	34*	34	-05	16	-07
33.	-42	-42	42	00	00	00	42	42	-100
34.	58*	21	27	50	27	58	--	100	-100
35.	00	-27	27	00	00	27	-27	27	00
36.	--	--	--	--	--	--	--	--	--
37.	-05	-05	13	05	34**	25	05	-13	34
38.	-02	-07	02	18	15	25	02	08	-05
39.	00	00	00	00	00	00	00	00	00
40.	-21	21	-21	-100	-21	-21	21	-21	27
41.	27	-27	25	14	25	25	-14	27	14
42.	-07	-34	20	34	34*	34	07	07	-07
43.	17	17	17	17	00	52	17	17	-34
44.	-27	-21	27	-21	21	21	21	21	27
45.	-52	-34	34	34	17	00	17	52	17
46.	-03	10	-10	-03	-07	-03	38***	50	-10
47.	-50	-50	-10	-10	10	-10	10	30	-30
48.	21	21	21	-21	21	-21	27	-21	27
49.	04	04	04	26	16	10	04	10	04
50.	07	-02	07	11	16	20	20	11	20
51.	09	12	12	09	20	16	-05	05	09
52.	08	02	08	17	52***	52	02	12	02
53.	30*	33	20	26	25	38	08	20	12
54.	24	24	40*	--	30	40	14	24	10
55.	00	07	-08	-03	08	11	31*	31	-03
56.	17	13	16	00	34*	30	03	-03	00
57.	07	07	07	14	14	18	14	21	00
58.	50**	20	08	-05	53***	53	-17	-05	-05
59.	26	02	14	20	48*	54	18	18	05
60.	11	-11	-04	08	44***	44	-11	-11	-16

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501	E100[1]
					WEAPONS EXAMINATION					
1.	16	26	26	20	20	26	57*	51	-08	-08
2.	08	03	14	19	14	19	03	-03	-06	00
3.	20	17	39**	36	39**	30	23	17	20	08
4.	02	02	-06	-06	-02	-02	22	02	40***	28
5.	00	42	-42	00	42	42	100	42	00	100
6.	00	00	00	00	00	00	00	00	00	36
7.	45	20	20	20	50*	60	-04	-20	04	11
8.	32*	60	-06	02	15	02	18	18	05	38
9.	-03	16	03	-03	10	10	26	22	03	24
10.	00	00	00	00	58*	56	00	00	00	-10
11.	40***	47	03	12	07	16	16	14	18	38
12.	02	-07	36	32	36	36	05	14	-05	13
13.	18	24	24	18	22	13	-03	10	22	18
14.	10	16	22	12	12	10	22	24	03	16
15.	10	10	07	14	-18	-05	37***	33	-03	05
16.	16	19	03	00	07	13	60***	60	03	20
17.	50***	30	09	30	24	30	00	09	24	48
18.	12	-07	-03	00	00	00	03	07	-18	07
19.	18	10	22	16	16	16	07	10	00	15
20.	00	00	25*	50	00	00	27	00	-27	-36
21.	14	02	14	23	26	23	11	-02	02	18
22.	16	16	-02	-03	13	03	00	12	-05	05
23.	00	13	28	28	16	10	24	13	07	18
24.	07	20	03	-10	07	03	07	10	03	-02
25.	03	12	29	18	21	10	18	10	21	15
26.	09	30	-16	-17	-09	-24	00	-22	16	22
27.	45**	38	26	15	52***	45	-20	-26	-03	06
28.	05	22	00	10	05	02	00	-05	05	02
29.	--	00	--	--	--	--	00	--	--	--
30.	05	20	16	24	16	23	03	08	03	24
31.	14	17	14	28	30***	46	02	12	08	02
32.	00	00	-03	-13	-13	-08	25**	-08	03	02
33.	15	20	-07	-08	12	12	17	12	20	16
34.	-05	05	05	11	02	10	24	20	02	05
35.	06	06	00	06	17	17	06	-06	00	04
36.	44	26	16	38	44	44	20	16	05	10
37.	21	28	28	41	32*	50	10	10	00	30
38.	-13	-16	47***	38	13	16	19	19	00	05
39.	07	10	20	17	20	23	32***	36	-05	-02
40.	14	18	10	18	10	18	10	10	06	12

[1] Cross Validation Sample

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501	E100
41.	22	22	29	22	29	40	00	17	00	10
42.	-05	-08	12	22	22	18	-08	02	05	12
43.	07	09	18	22	35*	44	05	05	02	09
44.	23	16	16	10	23	26	16	22	22	24
45.	-07	-12	07	20	20	26	00	00	-12	08
46.	09	06	16	16	12	24	-02	-02	08	-04
47.	24	14	33*	42	26	38	18	14	-02	02
48.	17	14	11	08	-02	05	11	22	08	02
49.	10	18	68***	58	32	37	18	18	-06	16
50.	-05	-05	52*	52	24	29	05	05	-10	38
51.	00	-08	37*	45	33	37	12	00	04	28
52.	22	10	34*	45	30	47	-10	06	10	33
53.	13	07	22	13	16	19	10	-07	62***	03
54.	13	17	20	27	20	27	20	20	-20	07
55.	-02	-02	-02	-06	16	06	46***	32	02	18
56.	04	11	11	36	08	08	20	15	00	20
57.	25	18	37	44	37	41	25	41	06	24
58.	22	13	29	38	29	32	07	22	00	03
59.	15	15	30*	20	24	38	10	05	05	17
60.	09	09	15	15	22	25	-12	-03	-05	18
61.	05	08	05	08	08	-03	-03	07	-12	13
62.	25	25	03	19	-03	03	-03	25	-08	40
63.	04	--	04	--	-04	04	--	--	--	-02
64.	00	05	26	20	32*	38	00	02	26	22
65.	07	00	03	07	10	-03	07	03	-07	-03
66.	30	35	35	38	25	25	35	32	-05	02
67.	00	07	11	11	11	23	07	23	07	06
68.	24	08	08	40	-40	24	24	-08	-24	20
69.	-10	02	50***	50	10	14	-02	06	02	30
70.	10	14	20	18	18	20	-07	07	03	14
71.	09	00	23	23	-05	19	00	05	05	14
72.	-10	-20	20	30	-07	07	-07	-07	00	20
73.	19	-05	13	09	23	27	-10	-13	-09	09
74.	16	10	31	22	31	40	-03	03	13	22
75.	10	10	10	24	20	15	10	15	00	00
				ENGINEERING EXAMINATION						
1.	13	16	03	00	07	07	34*	38	00	15
2.	-03	00	-07	-12	03	-03	-22	-26	-03	-07
3.	34	28	28	40	36*	36	12	05	08	20
4.	16	22	23	36	36*	44	23	23	12	24
5.	-08	08	08	08	17	08	17	30	02	05
6.	18	22	30	43	37*	33	07	16	30	-02
7.	-13	-10	-13	-14	-07	-10	-10	-22	-10	10
8.	17	17	07	-03	13	17	13	03	20	07
9.	06	06	-06	00	-06	-18	13	09	02	14
10.	16	08	16	32	28	32	-04	-02	-06	19

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501	E100
11.	18	18	22	33	22	18	33*	52	-06	08
12.	07	07	10	10	22	13	18	15	03	14
13.	28	16	45*	62	42	52	24	32	-06	11
14.	20	30	32	42	32	24	18	18	12	-02
15.	14	14	30*	30	26	14	-10	-10	06	-08
16.	14	10	07	-03	13	13	-22	-22	03	07
17.	14	12	26	24	26	28	24	08	30*	07
18.	12	06	44*	44	32	28	08	02	08	02
19.	28	28	44	36	44	41	-10	-02	25	21
20.	29	32	12	18	22	32	-02	-02	02	12
21.	-18	-05	-05	-05	-16	-12	-05	02	08	38
22.	39	10	28	39	39	28	03	07	07	-08
23.	00	-09	24	09	09	-06	04	09	16	30
24.	33	38	45*	62	38	42	18	22	-10	33
25.	08	02	35	32	42*	38	12	26	-05	13
26.	28	03	10	16	16	13	03	13	-10	-07
27.	10	00	03	-16	07	00	13	-03	16	17
28.	10	07	19	40	25	34	03	13	10	10
29.	40*	52	08	17	29	22	22	17	33	34
30.	26	20	26	32	26	20	14	20	08	19
31.	20	16	34	45	45*	42	28	24	-02	05
32.	-02	02	08	05	05	05	12	02	02	13
33.	19	06	02	12	06	10	-15	-06	-32	13
34.	25	25	16	25	29	20	12	08	04	11
35.	16	08	35	35	24	16	35	08	36	18
36.	24	20	30*	56	20	20	18	20	14	02
37.	13	13	34*	48	20	30	27	30	-03	31
38.	07	10	-07	-03	-16	13	10	16	13	-07
39.	06	10	15	17	19	10	06	19	-06	06
40.	00	03	10	20	27	23	10	07	-10	25
41.	22	30	30	71	50*	70	06	06	-06	36
42.	17	13	03	10	13	03	10	07	-07	28
43.	00	08	-07	-07	17	08	00	07	20	-03
44.	16	22	19	35	22	22	-03	07	03	41
45.	16	21	06	-03	12	08	08	08	-02	23
46.	13	13	13	19	19	13	09	27	05	09
47.	42	42	00	100	100	42	00	00	-100	100
48.	-02	11	07	11	-02	07	16	24	16	38
49.	14	24	07	03	20	20	10	10	-02	-15
50.	08	00	35*	30	20	23	20	23	-11	12
51.	35*	40	22	05	32	25	02	27	05	47
52.	18	15	15	30	15	20	08	24	-02	23
53.	16	23	38*	50	35	30	16	35	-16	17
54.	25	18	12	18	25	18	07	10	-16	-13
55.	04	26	15	45	20	20	27	15	-04	-04

TABLE 4.3 (continued)

Item No.	S105	E100	S207	E200	S309	E300	S401	E400	S501	E100
56.	32*	49	07	10	18	14	-03	-03	14	16
57.	11	08	10	15	28	30	20	18	06	07
58.	11	15	22	17	46*	52	29	33	-29	08
59.	-06	06	06	02	38*	34	-06	-06	18	22
60.	06	11	02	07	06	02	22	-06	15	20
61.	03	10	05	02	17	10	17	20	-07	-02
62.	22	12	02	15	12	15	02	12	-12	-18
63.	04	20	-20	09	-20	-20	20	12	-04	25
64.	-02	-02	02	05	05	05	12	12	-20	00
65.	00	00	00	00	00	00	00	00	00	100
66.	22	28	-22	-11	07	02	11	-07	07	22
67.	17	10	10	14	28	17	10	17	03	18
68.	11	22	22	40	30*	22	00	-11	06	17
69.	12	24	12	30	24	12	40*	49	-12	11
70.	42	21	21	21	21	21	-21	-42	66*	08
71.	-11	10	10	34	18	25	22	22	-08	-02
72.	08	24	13	13	28	28	13	13	13	08
73.	08	08	38*	38	29	18	12	12	08	08
74.	02	07	-16	-11	-02	-11	02	02	16	02
75.	08	05	12	16	18	16	18	12	08	12

At this point, further investigation of Keys S501 and E500 was abandoned, since these keys appeared to represent a residual factor. The reliabilities of scores on these keys were quite low, as will be shown later.

An analysis of the "overall" validities of the items on the E-keys (i.e., the validities against the composite score criterion described in Chapter 3) disclosed, however, that 60 (49.6%) of the 121 items on Keys E100, E200, F300, and E400 together included 57 per cent of the high validity items (r = .30 or greater) on the examination as a whole. The data of this analysis are shown in Table 4.4. It is to be noted that Keys E200 and E300 most frequently contained items with high overall validity; Keys E100 and E400 sampled from all ranges of overall item validity. It may be concluded that the dragnet operation of expanding the S-keys "caught" a large number of the available "good"items, and that at the same time some success had been achieved in identifying (at least statistically) aspects of achievement not represented very well, or masked, by the composite criterion. It is possible, however, that some of the items not caught by the dragnet may be measuring aspects of achievement which are not measured by the "E" keys. This possibility could have been investigated by Wherry-Gaylord iterative procedures, but time did not permit this project to do so.

The results were subjected to two types of checking procedures:

(1) Since the S-keys were not very reliable (as will be shown below), with a consequent attenuation of item-key correlations, the correlations of all items were computed against the E-keys. This was done for all E-keys except E500, which by this time the project staff was convinced was a residual factor not worthy of further investigation. The resulting correlations are shown in Table 4.3. In the main, it may be concluded from this table that the correlations of the items with the S-keys held up well when computed with the E-keys.

TABLE 4.4

FREQUENCY DISTRIBUTION - (OVERALL) VALIDITY COEFFICIENTS

Dec. '52, OCS Exam r	E100	E200	E300	E400	All Other Items	Total
.55-.59					2	2
.50-.54	2	1	1			4
.45-.49		2	4		4	10
.40-.44	1	4	7	1	2	15
.35-.39	1	10	10	3	16	40
.30-.34		4	6	3	21	34
.25-.29	3	4	7	1	48	63
.20-.24	2	5	3	4	49	63
.15-.19	4	1	5	5	41	56
.10-.14	2	2		3	35	42
.05-.09	4	1	2		25	32
.00-.04	2	1			24	27
-.05-.01					9	9
-.10-.06					10	10
-.15-.11					2	2
-.20-.16						0
-.25-.21						0
-.30-.26					1	1
TOTAL	21	35	45	20	289	410

(2) The item-key correlations were cross-validated on a new sample of 200 randomly selected cases, but only for items on the Weapons and the Engineering examinations, and only against key E100. Again, these results are displayed in appropriate places in Table 4.3. Some regression was apparent, but not enough to destroy the general pattern of the results. Data from the cross-validation sample will be discussed further, below.

Statistical Parameters of Scores on the Factor Keys

Table 4.5 shows the means, standard deviations, reliabilities, and homogeneity coefficients for the various factor keys in the sample used for the iterative factor analysis (Sample A) and the cross-validation sample (Sample B). Of particular concern were the reliabilities and homogeneity coefficients. The various factor keys were presumed to be fairly "pure" measures of certain aspects of achievement in Officer Candidate School, and they were presumed to be at least moderately reliable--more reliable (per item) than the original tests. Further, they were expected to result in higher homogeneity coefficients than the original full examinations.

The data of Table 4.5 must therefore be compared with those presented in Table 3.2 of the previous chapter. The following conclusions may be drawn:

(a) The S-keys, with the exception of Keys S401 and S501, maintained a higher average level of reliability-per-item than the original examinations. Part II of Navigation, which had a per-item reliability in Class VII of .107, failed to be surpassed by any of the S-keys, however. The same statement may be made with respect to homogeneity coefficients, that is, the S-keys in general surpassed the original examinations but failed to surpass Part II of Navigation.

TABLE 4.5

MEANS, STANDARD DEVIATIONS, RELIABILITY AND HOMOGENEITY COEFFICIENTS FOR FACTOR KEYS IN A STANDARDIZATION SAMPLE (A) AND A CROSS-VALIDATION SAMPLE (B)

N = 200 IN EACH SAMPLE

Key	Sample	No. of Items	$\bar{X}$	σ	K-R (20) rel.	r (1 item)	H_t
S105	A	17	12.16	2.04	.359	.032	.060
E100	A	21	14.45	2.42	.460	.039	.078
E100	B	21	14.64	2.36	.426	.034	.067
S207	A	13	8.05	2.31	.551	.086	.139
E200	A	35	24.72	4.68	.725	.070	.111
E200	B	35	23.86	4.24	.661	.053	.085
S309	A	23	16.18	3.03	.600	.061	.110
E300	A	45	30.63	5.59	.781	.073	.121
E300	B	45	30.08	4.71	.647	.039	.067
S401	A	15	8.87	1.99	.291	.027	.047
E400	A	20	12.91	2.80	.553	.058	.102
E400	B	20	12.70	2.39	.353	.044	.027
S501	A	11	7.72	1.38	.105	.011	.022
E500	A	15	10.02	1.77	.196	.016	.037

(b) In Sample A, the per-item reliabilities of the E-keys (excluding E500) clearly surpassed all the per-item reliabilities of the original examinations, except possibly Part II of Navigation.

(c) In Sample B, the cross-validation sample, the per-item-reliabilities and homogeneity coefficients showed a certain degree of regression or "shrinkage," as one might expect in a cross-validation sample. Nevertheless, they remained at levels which were still, in general, higher than the corresponding statistics in the complete tests.

Hence, the factor keys seem to measure certain aspects of achievement in a more reliable manner than the original department examinations did. If the factor tests could be increased in reliability by adding comparable material, the resulting tests would be more satisfactory measures of achievement than the original department examinations.

The above statements must be understood to apply to the first four S-keys and the corresponding E-keys. As noted previously, in view of the low reliability of S501 and E500, they were concluded to be "error factor" keys and not worthy of further investigation.

Correlational Data for Factor Keys

Before proceeding to attempt an interpretation of what aspects of achievement the factor keys may be presumed to measure, we shall present all available statistical data which may aid in this interpretation.

The question of greatest interest is whether the factors are truly independent, at least linearly independent. Do the factors measure different aspects of achievement? The most direct way of answering this question is to examine the intercorrelations among the E-keys, both in the standardization sample and in the cross-validation sample. The data are presented in Table 4.6 (the raw correlation coefficients) and Table 4.7 (the correlations corrected for attenuation).

Very little can be said about the results in Table 4.6, for the relatively low reliabilities of the E-keys obscure the pattern of relationships. In general,

TABLE 4.6

CORRELATIONS AMONG E-KEYS IN THE FACTOR ANALYSIS SAMPLE (A) AND THE CROSS-VALIDATION SAMPLE (B)

N = 200 IN EACH SAMPLE

Reliabilities are in the principal diagonal

Key	SAMPLE	E100	E200	E300	E400
E100	A	(.460)	.256	.305	[illegible]
	B	(.426)	.416	.374	[illegible]
E200	A	.256	(.725)	.612	.364
	B	.416	(.661)	.632	.478
E300	A	.305	.612	(.781)	.287
	B	.374	.632	(.647)	.512
E400	A	.252	.364	.287	(.553)
	B	.319	.478	.512	(.353)

TABLE 4.7

CORRELATIONS AMONG E-KEYS (TABLE 4.6) CORRECTED FOR ATTENUATION

Key	SAMPLE	E100	E200	E300	E400
E100	A	1.000	.444	.509	.500
	B	1.000	.783	.712	.822
E200	A	.444	1.000	.812	.575
	B	.783	1.000	.966	.988
E300	A	.509	.812	1.000	.437
	B	.711	.966	1.000	1.072
E400	A	.500	.575	.437	1.000
	B	.821	.988	1.072	1.000

the correlations are low, except between E200 and E300, which correlate to the extent of .612 in Sample A and .632 in Sample B. As might have been expected, the correlations in Sample B, the cross-validation sample, are in every case higher than in Sample A. This is because the iterative factor-analysis procedure capitalized on error in Sample A, at least to some extent, in producing sets of items which were relatively independent <u>in that sample</u>. In Sample B these errors have been again randomized, with the result that whatever inherent correlations exist among the items are manifested in higher inter-key correlations.

Table 4.7, containing correlations corrected for attenuation, presents a picture somewhat unfavorable to the hypothesis that the keys are linearly independent. The picture is not so unfavorable in Sample A; while keys E200 and E300 appear to be highly similar, keys E100 and E400 are somewhat independent of each other and of the remaining keys. (Of course, these data must be appraised in the light of the possibility that the estimates of reliability used in the correction for attenuation are in error.) But the picture is extremely unfavorable to the hypothesis of linear independence in Sample B. Here, all keys appear to have high inherent relations. Keys E200 and E300 have a corrected correlation of .966 and Key E400 correlates highly with all keys (its correlation of 1.072 with Key E400 is of course an overestimation). Only Key E100 maintains some degree of independence from the others. Again, however, the estimates of reliability may have been in error (i.e., too low, thus resulting in excessively high corrected correlations). This is particularly true for Key 400, the estimated reliability of which dropped to .353 in the cross-validation sample.

Our conclusion must be that while there is some evidence that the various E-keys are linearly independent (in the sense which is referred to in factor analysis), we must admit that there is a very strong general factor of achievement running through the E-keys. Certainly the keys are not as independent as was

hoped. From the standpoint of this evidence, grave doubt is cast upon the notion that it is possible to measure reliably different aspects of achievement at OCS.

Table 4.8 presents extensive correlational data for Class VIII (based on the iterative-factor analysis sample of 200 cases), giving the correlations within and among the following sets of variables:

(a) The Bureau examinations, by departments. (The correlations shown here represent sampling variations from those shown in one part of Table 3.5).

(b) Five S-keys, S105, S207, S309, S401, S501.

(c) Five corresponding E-keys, E100, E200, E300, E400, and E500.

(d) Three "F" keys, F1, F2, and F3. These were constructed from the corresponding E-keys for a special examination to be given to selected NROTC classes, and will be described and discussed in a later section of this report (Chapter VI).

(e) Five subtests of the Officer Candidate Battery (OCB): Verbal Reasoning, Mechanical Comprehension, Mathematical Comprehension, Relative Movement, and Spatial Relations.

We shall now discuss the findings which appear to be of most interest in these correlations.

(1) Correlations Between Bureau Examinations and S-keys.

It must be remembered, first of all, that in all cases of correlation between Bureau examinations and special keys, we are dealing with part-whole relations, since the special keys contained items from the Bureau examinations. Table 4.9 shows the numbers of items overlapping between Bureau examinations and the S-keys and the E-keys.

All examinations are more highly correlated with Key S207 and S309 than with the other two. Thus, there is no evidence that the keys differentially measure separate subject-matters. Taking the keys separately, furthermore, we find that their

TABLE 4.8

CORRELATIONS AMONG BUREAU EXAMINATIONS, SPECIAL FACTOR KEYS,

AND THE OFFICER CANDIDATE BATTERY

N = 200 (Factor Analysis Sample, OCS Class VIII)

All entries have been multiplied by 100 to eliminate decimal points.

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Bu Exam. Engin.	1	100	30	51	36	39	66	28	47	40	24	09	38	63	58	37	21	31	48	49	33	65	35	32	35
Bu Exam. Navig.	2	30	100	43	39	28	36	25	40	56	23	10	22	45	64	37	14	12	39	47	24	20	32	25	24
Bu Exam. Oper.	3	51	43	100	46	43	48	40	51	58	38	11	43	59	62	42	23	38	45	53	40	35	43	32	37
Bu Exam. Orient.	4	36	39	46	100	36	33	35	38	43	31	21	38	39	46	38	22	39	30	45	34	21	30	15	11
Bu Exam. Seam.	5	39	28	43	36	100	45	24	45	41	28	10	35	44	43	34	09	35	31	38	31	28	27	23	27
Bu Exam. Weap.	6	66	36	48	33	45	100	31	55	50	34	17	38	63	60	39	19	34	51	55	35	55	35	31	37
S105 Key	7	28	25	40	35	24	31	100	21	51	10	20	79	22	39	14	23	71	13	26	15	09	09	07	13
S207 Key	8	47	40	51	38	45	55	21	100	67	12	08	23	81	58	22	13	16	73	49	39	36	39	29	30
S309 Key	9	40	56	58	43	41	50	51	67	100	13	15	38	57	81	21	19	25	39	61	32	25	40	28	29
S401 Key	10	24	23	38	31	28	34	10	12	13	100	-01	20	22	17	83	12	21	17	16	15	17	26	16	18
S501 Key	11	09	10	11	21	10	17	20	08	15	-01	100	23	-06	13	-09	77	17	-04	13	03	07	05	04	02

TABLE. 4.8, continued

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
E100 Key	12	38	22	43	38	35	38	79	23	38	20	23	100	26	31	25	26	82	16	24	18	14	11	03	18
E200 Key	13	63	45	59	39	44	63	22	81	57	22	-06	26	100	61	36	05	19	82	55	42	43	45	36	35
E300 Key	14	58	64	62	46	43	60	39	58	81	17	13	31	61	100	29	18	21	46	86	40	46	44	38	39
E400 Key	15	37	37	42	38	34	39	14	22	21	83	-09	25	36	29	100	-03	24	26	26	21	26	31	18	17
E500 Key	16	21	14	23	22	09	19	23	13	19	12	77	26	05	18	-03	100	19	04	15	04	16	09	11	15
F 1 Key	17	31	12	38	39	35	34	71	16	25	21	17	82	19	21	24	19	100	10	20	15	04	01	-02	-12
F 2 Key	18	48	39	45	30	31	51	13	73	39	17	04	16	82	46	26	04	10	100	43	36	34	34	37	25
F 3 Key	19	49	47	53	45	38	55	26	49	61	16	13	24	55	86	26	15	20	43	100	42	42	38	32	36
OCB Verb. Reas.	20	33	24	40	34	31	35	15	39	32	15	03	18	42	40	21	04	15	36	42	100	28	40	21	22
OCB Mech. Comp.	21	65	20	35	21	28	55	09	36	25	17	07	14	43	46	26	16	04	34	42	28	100	46	35	49
OCB Math. Comp.	22	35	32	43	30	27	35	09	39	40	26	05	11	45	44	31	09	01	34	38	40	46	100	48	35
OCB Rel. Motion	23	32	25	32	15	23	31	07	29	28	16	04	03	36	38	18	11	-02	37	32	21	35	48	100	24
OCB Spat. Rel.	24	35	24	37	11	27	37	13	30	29	18	02	18	35	39	17	15	-12	25	36	22	49	35	24	100

TABLE 4.9

NUMBERS OF ITEMS ON BUREAU EXAMINATIONS
OVERLAPPING WITH FACTOR KEYS

Bureau Exam. (Dec. '52)	Total No. of items in Bureau Examination	Frequencies of items on Factor Keys							
		S105	S207	S309	S401	E100	E200	E300	E400
Engineering	75	0	0	0	0	3	9	9	3
Navigation	60	1	2	6	2	3	7	11	3
Operations	75	5	4	7	5	5	4	9	4
Orientation	75	7	2	5	3	6	3	7	5
Seamanship	50	1	1	2	0	1	3	2	0
Naval Weapons	75	3	4	3	5	3	9	7	5
TOTAL	410	17	13	23	15	21	35	45	20

correlations with the various examinations are approximately proportional. The correlations with the Operations examinations are most consistently the highest.

These results could be explained on either of two assumptions. Either the Bureau examinations are complex and contain different varieties of factor variance, or the four factor keys are really homogenous. The first of these assumptions seems somewhat preferable to the second.

(2) Correlations Between Bureau Examinations and E-keys.

Since the E-keys are merely expansions of the S-keys, it would be expected that the pattern of correlations between Bureau examinations and these keys would be similar, in most respects, to that with the S-keys. This expectation was confirmed by the data. Again, there was little evidence that the keys differentially measured different classifications of subject-matter.

(3) Correlations Between Bureau Examinations and OCB Scores.

Consideration of these correlations is somewhat outside the scope of this chapter, but in view of the interest which always attaches to data concerning the validity of aptitude tests, they will be discussed here. It is felt that the best way of analysing these data is in terms of the multiple correlations of the OCB scores with the several Bureau examinations. Table 4.10 presents the pertinent results.

It will be noted from Table 4.10 that the Engineering and the Weapons examinations are best predicted by the Mechanical Comprehension subtest of the OCB, with relatively little being added to the prediction by other subtests. These examinations admittedly include many items covering mechanical matters. The Navigation examination is best predicted by the Mathematical Comprehension subtest, with some contributions from Spatial Relations, Verbal Reasoning, and Relative Movement. The Operations, Orientation, and Seamanship examinations are consistently best predicted by Verbal Reasoning; this result accords with the conclusion one might reach by a careful inspection of these examinations, namely, that they involve, chiefly, verbal reasoning.

There are noticeable differences in the overall predictability of the examinations from the OCB. Engineering, Operations, and Weapons are predictable with multiple correlations from .552 to .670, while the remaining examinations have multiple correlations from .377 to .403.

(4) Correlations Between S-keys and E-keys.

These are in general only moderate, as would be desired for presumably independent factor keys, except for the high correlations between corresponding keys which are due to overlaps of items.

Since the S-keys were merely stepping stones to the expanded E-keys, there will be no further consideration of S-keys. We shall also omit mention of the F-keys, since these were developed only for a special purpose to be described later.

TABLE 4.10

MULTIPLE CORRELATIONS AND BETA WEIGHTS FOR THE PREDICTION OF BUREAU EXAMINATIONS FROM THE OFFICER CANDIDATE BATTERY

N = 200 in factor-analysis sample, OCS Class VIII

OCB Subtest	BUREAU EXAMINATIONS											
	ENGINEERING		NAVIGATION		OPERATIONS		ORIENTATION		SEAMANSHIP		WEAPONS	
	r	β	r	β	r	β	r	β	r	β	r	β
Verbal Reasoning	.329	.153	.245	.128	.403	.247	.342	.259	.308	.211	.353	.191
Mechanical Comprehension	.647	.570	.196	-.029	.352	.070	.214	.086	.278	.097	.546	.399
Mathematical Comprehension	.347	-.034	.315	.173	.426	.172	.298	.176	.269	.044	.354	.006
Relative Movement	.324	.099	.246	.112	.323	.119	.148	-.011	.233	.100	.314	.103
Spatial Relations	.352	.027	.244	.141	.366	.187	.114	-.047	.268	.133	.374	.109
Multiple R		.670		.377		.552		.390		.403		.600

(5) Correlations Between E-keys and OCB Scores.

If the E-keys truly measured separate aspects of achievement, it is conceivable that they might show different patterns of relations with a series of aptitude tests such as the Officer Candidate Battery. The Officer Candidate Battery consists of five tests designed to measure different aspects of ability. The correlations found among them and shown in Table 4.8 are only moderate and hence tend to support the notion that the tests measure different traits.

The technique of multiple correlation was again used to see whether the E-keys sorted out factors of achievement sufficiently well to be differentially predictable by the separate tests of the Officer Candidate Battery. The results are shown in Table 4.11.

It would appear that the E-keys are not (in general) differentially predictable from OCB subtests. To judge from the beta-weights, the first three keys are most weighted with Verbal Reasoning; at least, Verbal Reasoning makes the greatest unique contribution to the prediction of these keys. Key E100 is also weighted with a certain amount of Spatial Relations. Keys E200 and E300 are predicted to a substantial extent by all the subtests of the OCB, each subtest making some substantial amount of unique contribution to the variance of Keys E200 and E300. It could be argued from these results that Keys E200 and E300 represent "integrated knowledge" because they tap a number of different areas of ability, but this interpretation is somewhat dangerous (in view of the possible unreliability of the beta-weights) and involves a definition of "integrated knowledge" which was not envisaged at the time this project was designed.

In general, the results discussed above are fairly consistent with the psychological interpretations of the E-keys which will be described below. Key E400, however, presents a puzzle. To judge from the beta-weights in Table 4.11, it appears to be weighted most heavily with Mathematical Comprehension, whereas our psychological interpretation of the factor would seem to indicate that it measures simple memory for

TABLE 4.11

MULTIPLE CORRELATIONS AND BETA WEIGHTS FOR THE PREDICTION OF E-KEYS FROM THE OFFICER CANDIDATE BATTERY

N = 200 in factor analysis sample, OCS Class VIII

OCB Subtest	E100 r	E100 β	E200 r	E200 β	E300 r	E300 β	E400 r	E400 β
Verbal Reasoning	.178	.142	.416	.244	.397	.217	.208	.085
Mechanical Comprehension	.137	.048	.432	.184	.455	.195	.256	.127
Mathematical Comprehension	.109	.011	.451	.162	.445	.132	.306	.202
Relative Movement	.026	-.059	.360	.138	.381	.163	.176	.012
Spatial Relations	.177	.131	.348	.112	.393	.163	.172	.017
Multiple R		.234		.585		.600		.342

isolated facts. Nevertheless, the beta-weights on which the former interpretation is based are subject to sampling errors.

Psychological Interpretation of the E-keys

It has been seen, up to this point, that the evidence for differences among the achievement factors measured by the E-keys is rather slim. Furthermore, the E-keys do not show any important differences in their patterns of correlations with the Bureau examinations or with Officer Candidate Battery sub-tests. It is therefore only with extreme hesitation that an effort is made here to offer psychological interpretations of the E-keys.

Two members of the project staff examined the content of the E-keys with a view towards psychological interpretation. The approaches taken by these two persons were quite different, but led to somewhat similar conclusions.

Staff member A examined the data simply with the purpose of judging what mental processes the items of a given key required in common. His approach was

thus similar to that which is ordinarily taken in interpreting any results obtained by factor analysis. The interpretation made by this staff member is as follows:

Key E100. This key appears to represent a type of verbal reasoning in which judgments are made. The examinee is presented with a problem about which he makes a judgment, utilizing any information he may have, and at the same time using a variety of "common sense" in "figuring out" what the intended answer is on the basis of the structure of the alternatives presented. Very few of the items seem to require spatial, mechanical, or mathematical reasoning. Most of them do, however, involve factual information about the Navy and related subjects.

Key E200. This key appears to involve spatial and temporal reasoning, that is, reasoning in problems involving space, time, and sequences of events. Problems requiring mechanical reasoning are included in this definition.

Key E300. This key appears highly similar to Key E200, but tends to emphasize mathematical and numerical ability more than Key E200. It would be expected to correlate highly with Key E200. (This statement was made without knowledge of the actual correlation.)

Key E400. This key appears to involve nothing more than the memory for isolated bits of information, that is, factual information. If anything, one would expect the scholarly, bookish type of individual who likes to remember trivial details to get high scores on this key.

Staff member B decided to classify the items of the various keys with respect to the OCB subtests with which he thought each item would correlate most highly. This was done solely on the basis of the names of these subtests, since the actual test forms of the OCB were not available at the time this classification was done. Table 4.12 shows the classification made by this staff member.

It is apparent that the two staff members agreed that Keys E200 and E300 included larger numbers of mathematical and mechanical items than the remaining two

TABLE 4.12

SUBJECTIVE CLASSIFICATION BY A STAFF MEMBER OF ITEMS
OF THE E-KEYS WITH RESPECT TO OCB SUBTESTS

OCB Subtest	FREQUENCY			
	E100	E200	E300	E400
Verbal Reasoning	12	14	16	11
Mathematical Comprehension	5	10	11	2
Mechanical Comprehension	2	10	12	7
Spatial Relations and Relative Movement	2	1	6	0
TOTAL	21	35	45	20

keys. They also agreed that Key E100 measured verbal reasoning more than anything else. Beyond this, staff member B was handicapped in using the categories of the OCB. Using these categories, he could not have reached the conclusion that E400 measured simple memory for factual information; instead, he was forced to conclude that it chiefly measured verbal reasoning. On the whole, the two staff members agreed in their interpretations of the E-keys, a fact which lends some support to the notion that the E-keys are measuring different aspects of achievement in the officer candidate curriculum.

Concluding Statement

The research reported in this chapter attempted to sort out the items of the Bureau OCS examinations into a number of sets measuring different aspects of achievement. The research was only partially successful in attaining this objective. The sets of items, while appearing to be different from a subjective point of view, were not sufficiently independent from a statistical point of view.

The question may now be raised: what do the results show with regard to "integrated knowledge"? It may be suggested that no one of the E-keys measures

anything distinctly recognizable as "integrated knowledge"; instead, all of them do, at least keys E100, E200, and E300. Only key E400 appears (at least to one staff member) to measure simple factual information. Keys E200 and E300 (which are in any case highly correlated) appear to measure the technical and mathematical kinds of subject matter which are so predominant in the officer candidate curriculum. Key E100, however, appears to measure something slightly different, namely, judgmental problem-solving. This aspect of achievement is undoubtedly of interest to the Navy.

It is recommended that the E-keys would be of value in future research in at least two respects:

(1) as sets of items against which new items could be validated; and

(2) as predictors of later success in the Navy.

It is conceivable, for example, that key E100 measures those aspects of problem-solving which are most required in the novel situations which the naval officer is constantly encountering.

For either of these purposes, it might be well to expand the E-keys even further. That is, items with validities against the E-keys somewhat lower than .30 (say, down to .20) could be added to the keys. Reference may be had to Table 4.3 in selecting such items. Items should, however, be added to the E-keys in such a way that there is no overlap among the newly-expanded keys.

CHAPTER V

THE TECHNIQUE OF OPERATING CHARACTERISTICS

Introduction

As an introduction to this chapter, it may be pertinent to quote from a chapter found in a recent textbook on educational measurement:*

"Most educational achievement tests are concerned primarily with power or level rather than rate. There is some theoretical basis for contending that the performance on a power test should be expressed not in terms of raw scores, but in terms of the level of difficulty of the most difficult group of items to which the examinee is able to respond correctly some set proportion of the time, such as 50 percent. The test might consist, for example, of ten sets of items grouped according to difficulty, each successive set being more difficult than the set preceding. An individual's score on this test may be 7.2, meaning that he is able to respond correctly to approximately half of the items in the set that has been given a scale value of 7.2, but that he responds correctly to fewer than half of the items in the higher sets and to more than half in the lower. There are a number of very serious practical and theoretical difficulties in constructing such tests, one of the most troublesome being that the items must be highly homogeneous, or highly correlated with one another. Variations in the learning experiences of different pupils are such that few types of items can be expected to be homogeneous, or to maintain a stable comparative difficulty. For this and other reasons, tests reporting performance in terms of level of difficulty are rarely constructed in educational measurement."

Despite the difficulties pointed out in this quotation, it is believed that a technique of scaling scores in terms of item difficulties has not been given the attention it deserves. It is also believed that the results to be achieved can have important practical implications for the interpretation and use of tests. In the case of the present research, use of this technique can make results more meaningful to the Navy. This chapter, therefore, will sketch the theoretical basis for the technique and will report and interpret some results attained by its use.

---

* Lindquist, E. F. (editor) Educational Measurement. Washington, D.C.: American Council on Education, 1951. This quotation is from page 706.

Theoretical Basis

It is commonly assumed that traits measured by psychological and educational tests exist along continua, that is, that the traits exist in various amounts in different individuals. For any given trait, therefore, we may represent the amount of that trait along the baseline of a graph.

In the case of tests of ability which involve the "passing" or "failing" of items, we may also be concerned with the relation between the amount of the trait which an individual possesses and the probability of his passing items at various levels of difficulty. We may assume that individuals with high ability will pass items of high difficulty with a higher probability than an individual of low ability.

If we further assume that all items in a given set measure the same trait, we may arrange these items along the scale of ability mentioned earlier, so that the items themselves define various amounts of ability. Thus, high-difficulty items define the upper reaches of the scale of ability in the sense that high-ability individuals are able to pass them with some specified probability (e.g. .5). Likewise, low-difficulty items define the lower reaches of the scale of ability in the sense that even individuals of low ability are able to pass them with the specified probability.

In practice, the best estimate we can make of an individual's ability is in terms of his score (e.g., number right) on a set of items. This score has the function of summating the probabilities with which he can pass the various items in a set. In fact, the score is a direct function of the average probability with which he can pass items in the set. If, however, we determine or estimate in some way the probabilities with which an individual with a given score can pass items of various difficulties, we can make the meaning of this score more apparent. Graphically, we may draw a curve representing these probabilities. Such a curve is called an <u>individual operating characteristic</u> (not to be confused with the <u>item</u> operating characteristics studied by

Lord.)* Figure 5.1 illustrates an idealized individual operating characteristic. The curve is to be interpreted as showing that an individual with a certain score on a test (say, 20) will pass items of 90% difficulty (for the group) with a probability of .005; items of 50% difficulty, with a probability of .25; item of 10% difficulty, with a probability of .880; etc. For the purposes of graphical presentation, some shape for the operating characteristic curve has to be assumed. There is good reason to assume, in practice, that operating characteristic curves have the shape of normal ogives.

It is, however, possible to conceive a type of test, or a type of trait, that would have what we shall call a perfect operating characteristic. A perfect individual operating characteristic is illustrated in Figure 5.2, and may be interpreted by stating that an individual with a certain degree of ability will pass all items up to a certain level of difficulty, but will fail all items beyond that level of difficulty. If all individual operating characteristics are of this shape, the test would be precisely what Loevinger demands for a "perfectly homogeneous test." In fact, a perfectly homogeneous test absolutely implies that all operating characteristics be perfect.** It is also true of this type of test that all items would have tetrachoric intercorrelations of unity.

It is, however, inconceivable that tests or traits having perfect operating characteristics would ever be found in practice. Even in the most carefully controlled psychophysical experiments, the psychophysical curves (which are analogous to our individual operating characteristic curves) do not have the shape shown in

---

* Lord, F. M. A theory of test scores. Psychometric Monographs, No. 7.
** Or nearly so. The requirement is actually that the individual operating characteristic curves always move from unity at a given level of difficulty to zero for the item of the next higher level of difficulty. Thus, the central portion of the operating characteristic curves could depart slightly from a perfectly vertical position when two adjacent items are slightly different in difficulty.

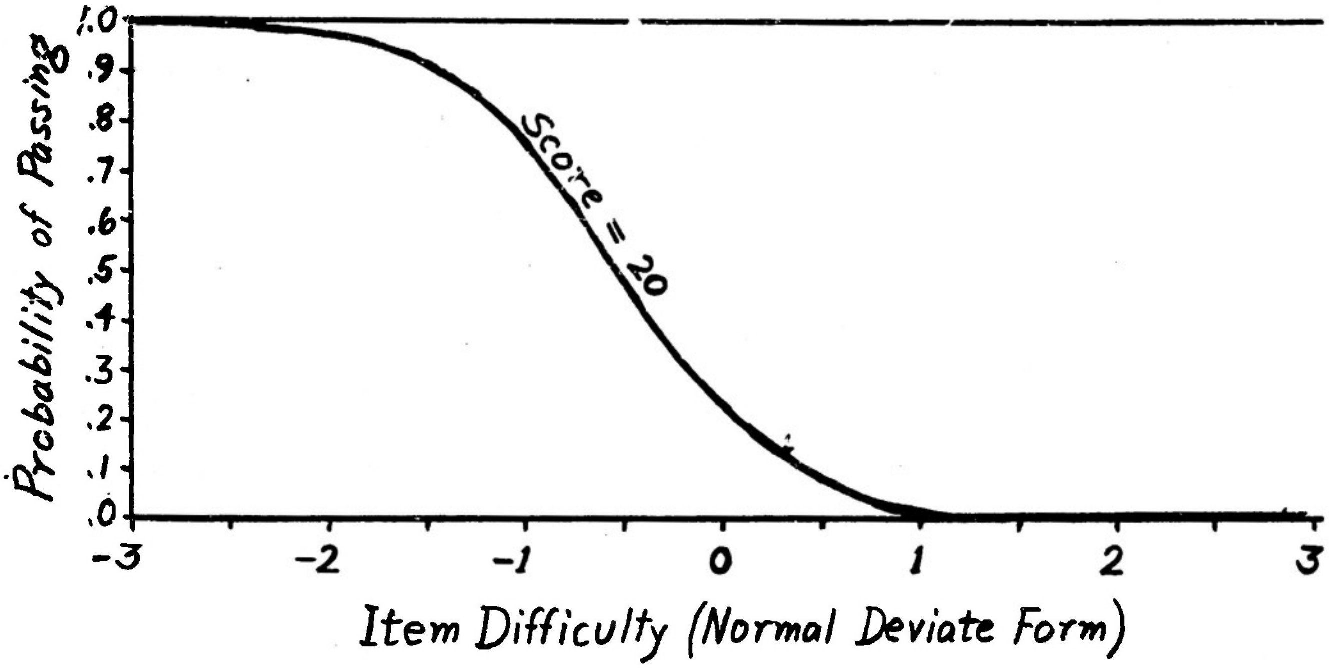

Figure 5.1. Operating Characteristic Curve For an Individual with a Score of 20. (Hypothetical Data) The curve shows the probability which the individual has of passing items of any given degree of difficulty.

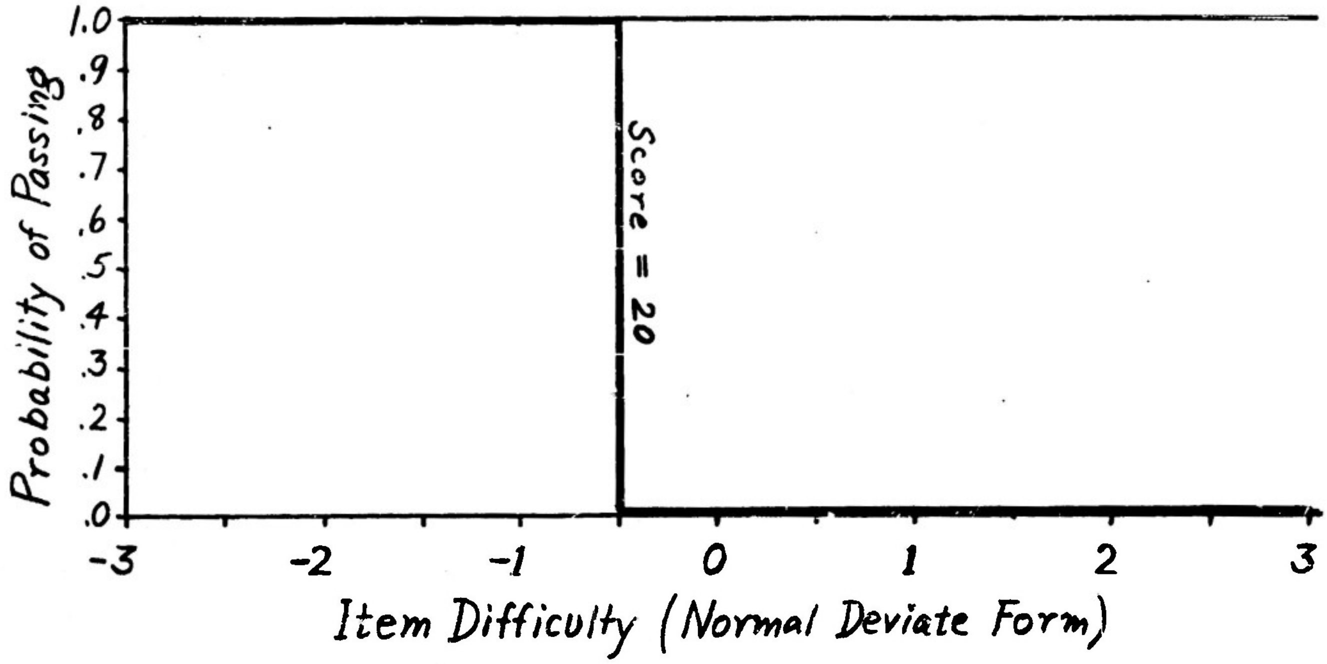

Figure 5.2. A perfect operating characteristic curve. The individual's score is still 20.

Figure 5.2. Furthermore, perfect operating characteristic curves would not be found in the case of mechanical or electronic equipment (e.g., a vacuum tube in a relay circuit). It therefore seems reasonable to conclude that the perfect operating characteristic represents an impossible ideal, and that meaningful results can be attained with tests having imperfect operating characteristic curves. The quotation from Lindquist cited above seems to imply that the scaling of scores on achievement tests with respect to item difficulties cannot be done unless the operating characteristics are perfect or nearly perfect; the present proposal assumes, however, that meaningful scaling can be done even when the operating characteristics are imperfect, that is, when the slopes are not vertical but assume the shape of something like a normal ogive.

The technique we shall employ will involve the estimation of probabilities-of-passing for various test scores and for items of various difficulties. Instead of using individual operating characteristic curves, however, we shall find average individual operating characteristic (O.C.) curves for all individuals making a specified observed score on the test. Furthermore, we shall find the probabilities of passing for groups of items of similar difficulties, in view of the fact that a very high N would be required to estimate the probabilities for individual items with precision. The average O. C. curves will be plotted on a probability graph, the item difficulties being scaled on the abscissa in terms of their normal deviate values, and the probabilities of passing being plotted on the ordinate, likewise in terms of their normal deviate values. In this way, the plot will reveal the extent to which the O. C. curves conform to the assumption that they have the shape of normal ogives, and further, the extent to which they have identical slopes. If all averaged individual O. C. curves have the same normal ogive shape and slope, the resulting plot will look like Figure 5.3.

Some question might be raised about the use of the observed score as the basis for an O. C. curve. It would, of course, be better to use true scores. However, the true scores are obviously inaccessible, and even "regressed" scores (as estimates of

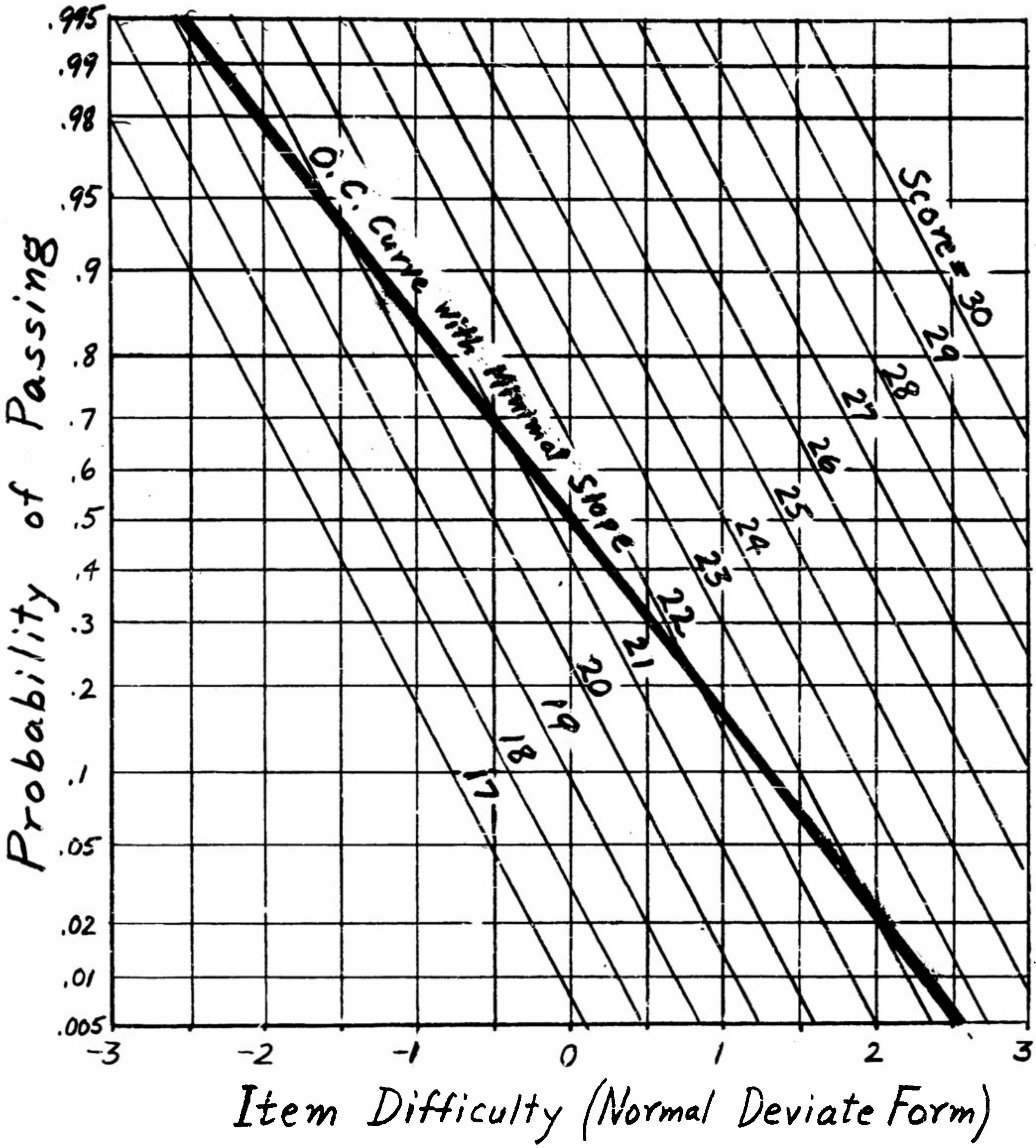

Figure 5.3. Operating Characteristic Curves for various score levels on a hypothetical 50 item test.

the true score) are simply linear functions of the observed scores. Hence, the observed scores can be used with justification.

The slope of the O. C. curve when plotted on probability paper (i.e. with both coordinates in normal deviate form) should presumably have something to do with the homogeneity of a test. This matter has not been thoroughly investigated, due to lack of time, either from an empirical or a mathematical standpoint. It is obvious, however, that the O. C. curves for a perfectly homogeneous test would appear as perfectly vertical lines of the O. C. plot on probability paper. Further, for a perfectly heterogeneous test, the variance of scores would be solely error variance, and insofar as the items might vary in difficulty the O. C. curves would parallel the curve for the average individual in the total group. Thus, the minimal operating characteristic would be represented on the O. C. plot with a 45° slope such that the normal deviate value of the baseline would equal the normal deviate value of the abscissa. An O. C. with this minimal slope has been plotted on Figure 5.3. The homogeneity of a test is therefore a function of the extent to which its O. C. curves tend to tilt toward the vertical, measuring from the minimal 45° position. As a tentative measure of homogeneity or operating characteristic, one might suggest the measure

$$\frac{\alpha - 45^{\circ}}{45^{\circ}}$$

where α is the average angle of the slope of the O. C. curves. For the investigation of the slopes of the O. C. curves which will be presented below, it would be necessary to apply or devise some sort of curve-fitting technique, e.g. some technique akin to those used in psychophysics, in order to find the average angle of slope, α . This procedure would undoubtedly be extremely laborious, however. The estimates of reliability and homogeneity made elsewhere in this report are regarded as sufficiently useful to make unnecessary the laborious investigation

of O. C. slopes which might otherwise be made. It will be apparent, however, that the operating characteristics of most of the curves which will be shown here are extremely low. This fact does not vitiate the utility of the technique of operating characteristics.

<u>Importance for the Navy</u>

The importance of this technique for the Navy may now be stated. The scores on nearly all educational and psychological tests are usually interpreted in terms of relative, group standards. For example, a score on a certain test might be said to have a percentile rank of 65, or a Navy standard score of 54, or a stanine of 6, etc., but always with respect to the actual ability of the examinee: it does not specify the kinds of items the examinee can pass, nor the probabilities with which he will pass them.

Suppose, therefore, one is interested in establishing a critical score such that students below that score will not be considered to have passed a naval officer candidate course. If one uses only relative standards, the decision becomes purely arbitrary. With the use of O. C. curves, however, one can find that score which seems to stand at the critical point between acceptable and unacceptable candidates, <u>in terms of what these candidates know or can do</u>. For this purpose, the O. C. graphs must be used in conjunction with the actual items on the test, preferably ranked in difficulty to correspond to the scaling of the items on the baseline of the O. C. plot.

It is suggested that the items be inspected with a view to finding the point on the item difficulty scale where it is desired that the minimally acceptable candidate will be able to pass the items with some specified probability, say .75 (that is, 3 times out of 4, or 75% of the time). The O. C. graph can then be used to find the test score which corresponds to this requirement. For example, suppose after inspection of the actual items represented along the baseline of Figure 5.3, it is decided

that minimally acceptable candidates should be able to pass, 75% of the time, an item with a scale value of -.64. It will be seen from Figure 5.3 that this requirement corresponds most closely to a score of 21. By reference to the distribution of scores, the percentage attaining this critical score in any specified group can, of course, be determined, but it is believed more meaningful to <u>first</u> determine the critical score by reference to the items rather than by reference to the percentage attaining a given critical score.

The utility of the technique of O. C. curves in comparing groups is also apparent. Once a critical score is established (<u>without</u> reference to the distribution of scores) the percentages attaining the critical score in various groups can be determined and compared. This procedure will be illustrated in Chapter 6 of this report, where OCS and NRCTC groups are compared.

<u>Operating Characteristic Curves for E-keys in OCS Class VIII</u>

The technique of operating characteristics is probably not appropriate unless the items analyzed have at least some degree of homogeneity. It is believed that the iterative factor-analysis techniques employed in the analysis of the OCS examinations produced sets of items (the "E-keys") sufficiently homogeneous to justify the use of O. C. analysis. It will be recalled that each item in the E-keys had a tetrachoric correlation of .30 or greater with its respective E-key. This insures that <u>item</u> O. C.'s have at least moderate slopes and hence in general insures the separation of individual O. C.'s.*

---

* An <u>item</u> O. C. represents the relation between ability (as estimated by the <u>observed</u> score) and the proportions passing a <u>given</u> item. Thus, the <u>item</u> O. C. curve is represented by the intercepts of <u>individual</u> O. C.'s with the perpendicular at a given point on the baseline of our O. C. plots.

The O. C. plots for keys E100, E200, E300, and E400 as applied to OCS Class VIII examinations are shown in Figures 5.4, 5.5, 5.6, and 5.7 respectively. Tables 5.1 to 5.4 give the data on which the curves are based and identify the items scaled along the base-line. The total number of cases for each plot is 200; the iterative factor-analysis sample was used.

Since the use of these plots involves reference to the actual items, which are under a security classification, it is not possible to discuss the implications of these plots beyond pointing out that the O. C. curves conform reasonably well to the hypothesis that they are normal ogives (i.e. straight lines when plotted on normal deviate coordinates). The parallelism among the O. C. curves for the various scores should be pointed out.

It may be concluded that a test score is shown to have high validity in predicting the <u>probability</u> with which an individual will pass an item of a given difficulty, that the relationship between item difficulty and probability of passing can be expressed as a linear function, and that the O. C. curves form a family of curves whose separation can be described in terms of a single parameter.

The occasional instances where the O. C. curves cross occur chiefly at the extremes of difficulty, where the estimates of probability of passing are undoubtedly less reliable. An effort was made to make them as reliable as possible by basing them on reasonably large numbers of cases. This was accomplished by grouping scores toward the ends of the distribution.

As was pointed out earlier, the operating characteristics found for these keys are extremely low, in the sense that their slopes are almost insensibly greater than the 45° slope which would obtain for a perfectly heterogeneous test. It will be remembered that the homogeneity coefficients computed for these keys and reported in Chapter 4 are uniformly low. This means that the tetrachoric correlations among the items are low.

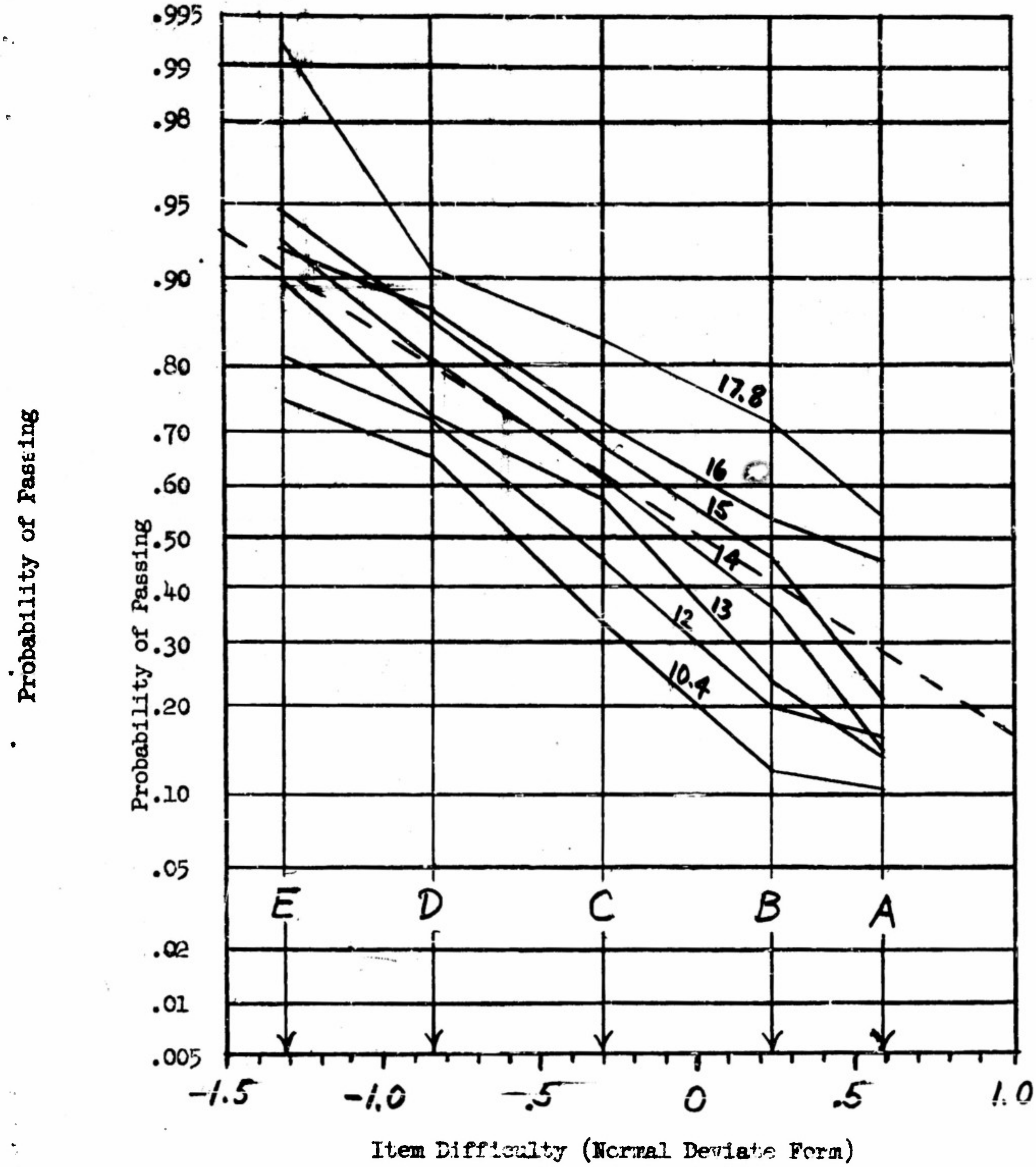

Figure 5.4. Averaged individual O. C.'s for various score groups on Key E100. The broken line extending diagonally across the whole plot represents the expected minimal slope of the O. C. curve for a perfectly heterogeneous test. The data for this figure are given in Table 5.1.

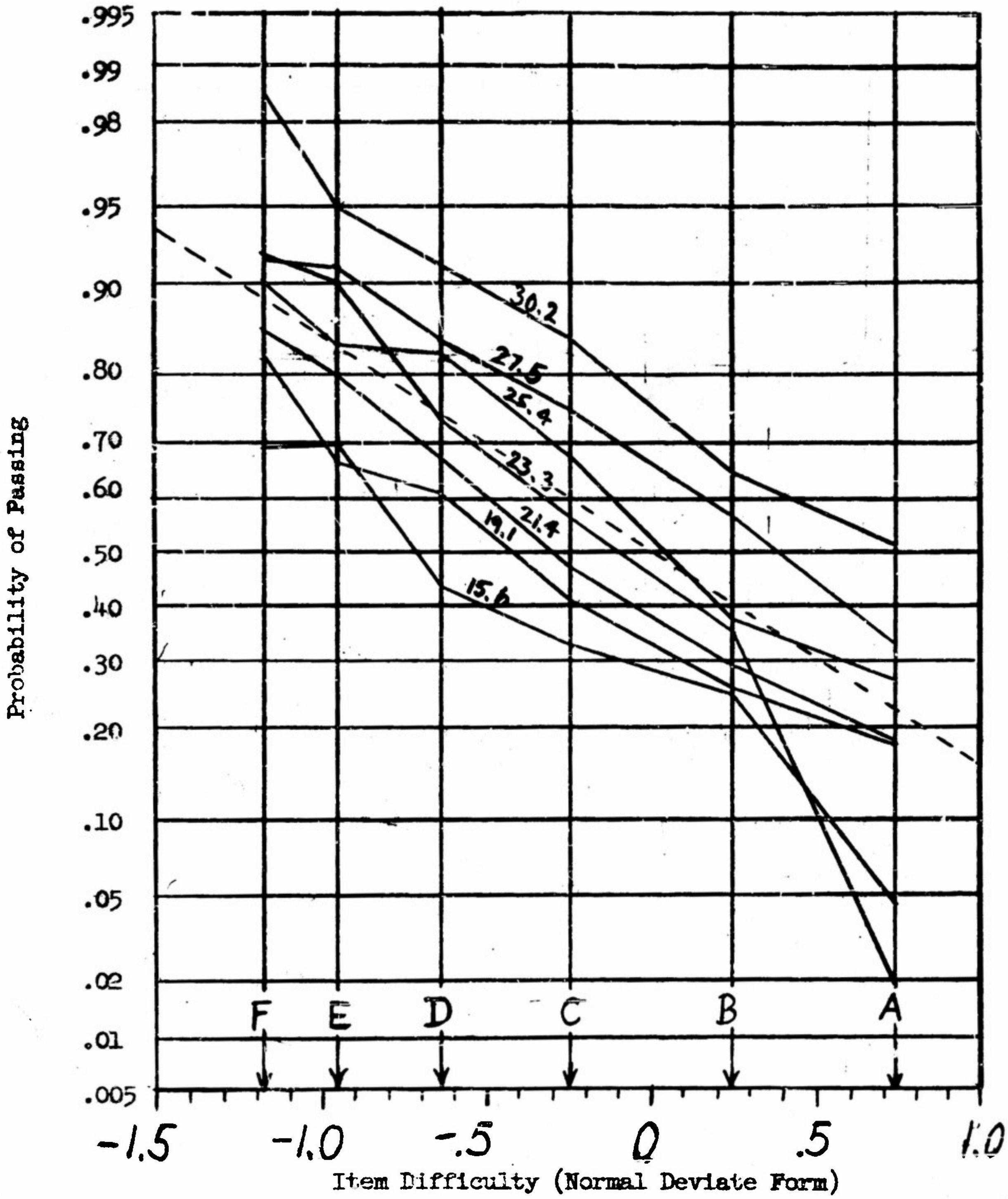

Figure 5.5. Averaged individual O. C.'s for various score groups on Key E200. The broken line extending diagonally across the whole plot represents the expected minimal slope of the O. C. curve for a perfectly heterogeneous test. The data for this figure are given in Table 5.2.

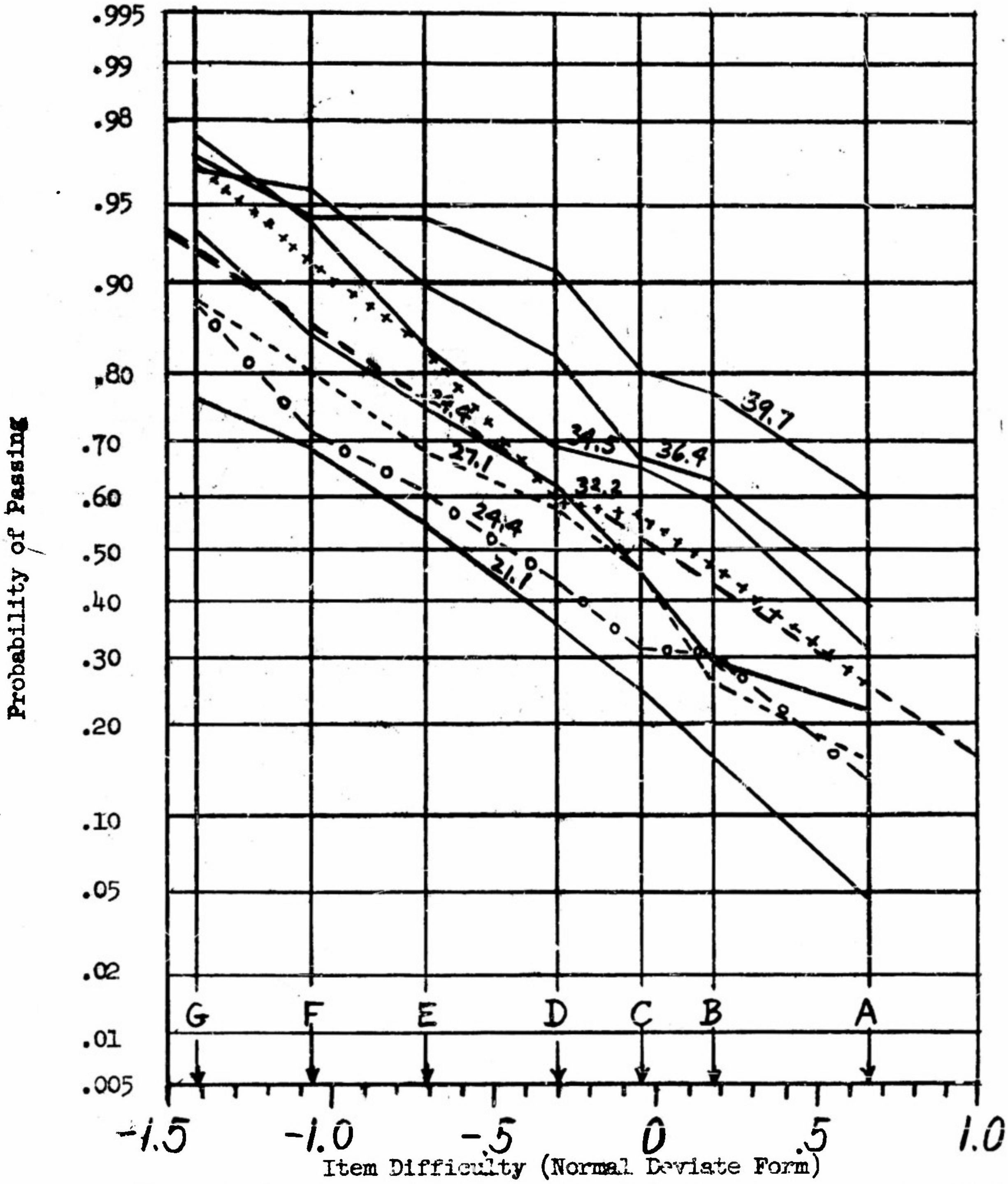

Figure 5.6. Averaged individual O. C.'s for various score groups on Key E300. The broken line extending diagonally across the whole plot represents the expected minimal slope of the O. C. curve for a perfectly heterogeneous test. The data for this figure are given in Table 5.3.

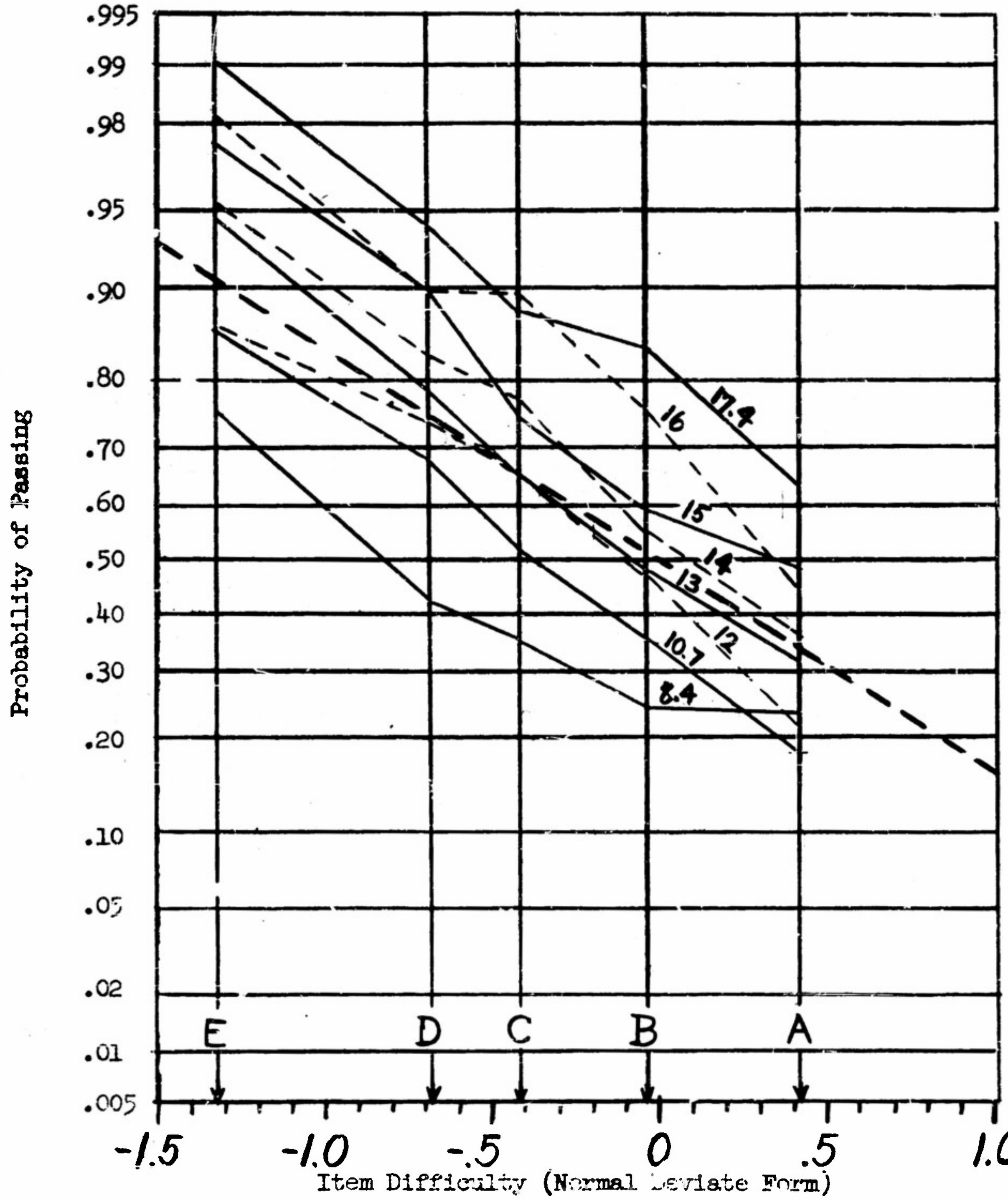

Figure 5.7. Averaged individual O.C.'s for various score groups on Key E400. The broken line extending diagonally across the whole plot represents the expected minimal slope of the O. C. curve for a perfectly heterogeneous test. The data for this figure are given in Table 5.4.

TABLE 5.1

DATA FOR OPERATING CHARACTERISTIC CURVES, KEY E100,

OCS CLASS VIII EXAMINATIONS

Total N = 200 cases in iterative factor-analysis sample

Item Group	Mean normal deviate value	Proportion Passing Score (Rights) 7-11 ($\bar{X}$=10.4)	12	13	14	15	16	17-19 ($\bar{X}$=17.8)
A	.59	.107	.160	.136	.145	.208	.452	.542
B	.23	.121	.200	.234	.362	.458	.538	.708
C	-.31	.333	.453	.568	.609	.667	.710	.834
D	-.86	.648	.720	.722	.804	.859	.871	.906
E	-1.31	.747	.811	.899	.925	.947	.922	.993
No. of cases in Score group:		22	25	27	23	32	31	40

SPECIFICATION OF ITEM GROUPS

(The item numbers on the respective examination and the total proportion passing are given)

ITEM GROUP

A #	A p	B #	B p	C #	C p	D #	D p	E* #	E* p
N-53	.230	W-8	.345	O-21	.600	Op-28	.780	E-29	.860
E-56	.300	W-11	.430	E-51	.625	S-26	.805	Op-69	.890
Op-70	.305	O-51	.455	O-40	.645	O-20	.815	W-17	.910
						Op-75	.820	N-24	.910
								O-65	.915
								Op-71	.920
								O-5	.925

*N-34, with a p of .970, was omitted from this group.

TABLE 5.2

DATA FOR OPERATING CHARACTERISTIC CURVES, KEY E200, OCS CLASS VIII EXAMINATIONS

Total N = 200 cases in iterative factor-analysis sample

Item Group	Mean normal deviate value	Proportion Passing — Score Groups (Range and Mean) (Rights): 12-17 / 15.6	18-20 / 19.1	21-22 / 21.4	23-24 / 23.3	25-26 / 25.4	27-28 / 27.5	29-34 / 30.2
A	.73	.046	.183	.185	.018	.267	.333	.514
B	.24	.250	.258	.296	.357	.375	.574	.646
C	-.24	.325	.410	.471	.561	.671	.751	.849
D	-.63	.428	.608	.676	.732	.828	.840	.917
E	-.95	.700	.660	.800	.900	.833	.911	.950
F	-1.18	.693	.825	.852	.920	.900	.917	.986
No. of cases in score group:		22	30	27	28	30	27	36

SPECIFICATION OF ITEM GROUPS

(The item numbers on the respective examinations and the total proportion passing are given)

ITEM GROUP

A #	A p	B #	B p	C #	C p	D #	D p	E #	E p	F* #	F* p
N-54	.180	O-9	.370	E-53	.515	N-15	.690	W-50	.820	Op-13	.850
Op-40	.295	W-38	.410	W-47	.565	Op-25	.700	N-20	.825	N-5	.865
		O-46	.415	S-30	.565	N-25	.710	W-59	.830	S-18	.890
		O-58	.430	E-36	.585	E-13	.715	Op-20	.835	N-17	.910
				S-40	.600	E-18	.725	E-73	.835		
				E-37	.660	W-49	.735				
				N-16	.670	E-50	.740				
						E-24	.745				
						W-52	.755				
						W-69	.765				
						E-15	.765				
						W-51	.770				

N = Navigation
W = Naval Weapons
Op = Operations
O = Orientation
E = Engineering
S = Seamanship

*W-20, with a p of .970 was omitted from this group.

TABLE 5.3

DATA FOR OPERATING CHARACTERISTIC CURVES, KEY E300,

OCS CLASS VIII EXAMINATIONS

Total N = 200 cases in iterative factor-analysis sample

Item Group	Mean normal deviate value	Proportion Passing — Score Groups (Range and Mean) (Rights) 17-23 21.1	24-25 24.4	26-28 27.1	29-30 29.4	31-33 32.2	34-35 34.5	36-37 36.4	38-43 39.7
A	.66	.045	.136	.153	.218	.248	.307	.391	.600
B	.18	.164	.300	.258	.290	.469	.584	.627	.770
C	-.04	.247	.319	.452	.458	.563	.651	.665	.800
D	-.30	.356	.439	.576	.615	.590	.693	.819	.908
E	-.71	.545	.608	.677	.746	.832	.830	.897	.944
F	-1.07	.692	.707	.801	.854	.914	.942	.957	.944
G	-1.41	.764	.882	.883	.938	.966	.976	.965	.970
No. of cases in score group:		22	22	24	29	35	25	23	20

SPECIFICATION OF ITEM GROUPS

(The item numbers on the respective examinations and the total proportion passing are given)

ITEM GROUP

A #	A p	B #	B p	C #	C p	D #	D p	E #	E p	F #	F p	G* #	G* p
Op-26	.235	W-31	.405	E-4	.505	Op-17	.585	E-31	.705	N-52	.825	W-7	.905
N-60	.240	N-59	.415	Op-30	.510	E-3	.585	Op-11	.715	0-70	.830	E-41	.915
E-6	.290	Op-31	.420	W-64	.515	N-29	.595	W-37	.735	0-16	.835	E-68	.915
		Op-6	.420	E-25	.515	S-13	.635	E-59	.755	0-4	.850	Op-21	.930
		N-58	.485	S-4	.525	0-72	.655	N-4	.784	E-58	.860	N-42	.935
				N-7	.530	N-56	.660	Op-8	.785	W-27	.865		
				N-32	.530			Op-12	.795	N-3	.880		
								W-43	.800	0-28	.880		
										0-15	.885		

* 0-56 and W-10, with p's of .960 and .970, respectively, were omitted from this group.

TABLE 5.4

DATA FOR OPERATING CHARACTERISTIC CURVES, KEY E400,

OCS CLASS VIII EXAMINATIONS

Total N = 200 cases -- iterative factor-analysis sample

Item Group	Mean normal deviate value	Proportion Passing — Score Groups (Range and Mean) (Rights)							
		6-9 8.4	10-11 10.7	12 12.0	13 13.0	14 14.0	15 15.0	16 16.0	17-20 17.4
A	.42	.232	.179	.214	.315	.360	.485	.443	.636
B	-.03	.241	.350	.473	.478	.550	.589	.761	.841
C	-.41	.345	.514	.655	.652	.773	.745	.894	.879
D	-.68	.420	.679	.741	.783	.830	.897	.898	.943
E	-1.32	.757	.857	.864	.948	.952	.976	.982	.991
No. of cases in score group:		28	35	28	23	25	17	22	22

SPECIFICATION OF ITEM GROUPS

(The item numbers on the respective examination and the total proportion passing are given)

ITEM GROUP									
A		B		C		D		E	
#	p	#	p	#	p	#	p	#	p
W-15	.280	O-63	.445	Op-54	.615	W-55	.725	W-1	.865
Op-60	.325	N-55	.460	Op-1	.655	E-11	.745	O-53	.900
O-31	.360	W-16	.570	N-46	.705	N-18	.760	O-8	.905
W-39	.380	E-1	.570			O-6	.780	Op-2	.925
								E-69	.930

CHAPTER VI

TECHNIQUES FOR THE COMPARISON OF NROTC AND OCS CANDIDATES

Introduction

It was pointed out in Chapter I that the curricula for NROTC and OCS candidates differ to a considerable extent--certainly in arrangement, and to some extent in slant and content. The examinations for the two programs must of necessity be different, for they must reflect the curricular differences. Therefore, if it is desired to compare graduates of the two programs with respect to their achievement in naval science, it is impossible to use the examinations as they stand. This chapter is concerned with an attempt to adapt the examinations in such a way that reasonable comparisons between NROTC and OCS candidates could be made. It is believed that the techniques developed here could also be adapted for use in comparing ROC (Reserve Officer Candidate) graduates with the other two groups.

How well do 4th year NROTC students retain the knowledge acquired during the first three years?

One factor which militates against a valid comparison of NROTC and OCS graduates is the fact that in NROTC the subject-matter is introduced piecemeal over a four-year period, whereas in OCS six different (but somewhat overlapping) subject-matters are taught together in the course of 16 weeks. At the end of the OCS course, the student is presumably prepared to be examined on any subject matter, while the 4th year NROTC graduate is best prepared only in Engineering, certain aspects of Military Justice, and several small parts of other subject-matters.

In order to determine how well fourth-year NROTC graduates retain the material of the first three years, a special 54-item examination, hereafter called the Cumulative Examination, was constructed for administration to fourth-year students at certain NROTC units in May 1953. This examination was composed chiefly of items from keys E100, E200, and E300 of the OCS examinations which appeared to be appropriate for testing subject-matter covered in the NROTC curriculum of the first three years. Table 6.1 identifies the items on this examination. The items were classified in three keys, F1, F2, and F3, corresponding to Key E100, E200, and E300, respectively.

It had been hoped that it might be possible to administer the NROTC Cumulative Examination to a random sample of NROTC graduates, using various NROTC units across the country. This was not feasible, however. Circumstances made it possible to give the examination to fourth-year students at only four NROTC units, those at Harvard University (N = 41), Yale University (N = 50), Brown University (N = 36), and the College of the Holy Cross, Worcester, Massachusetts (N = 29), a total of 156 students. Since it was thought that these four units might comprise a biased sample of NROTC students, scores on the OCB were obtained for the students tested, as a means of statistical control on their ability level.

Next, the test papers of the item-analysis sample (N = 200) of OCS Class VIII cases were scored with keys F1, F2, and F3. (It will be recalled that these scores were involved in the correlation matrix presented in Table 4.8.) The OCB scores were, of course, also available for this sample.

Table 6.2 shows the means, standard deviations, and reliability and homogeneity coefficients for the F-keys of the cumulative examination in the OCS sample and in the NROTC sample. Two major conclusions may be drawn from this table. First, it is evident that the NROTC group did less well than the OCS group. This result was expected, because the NROTC students were tested on materials they had not reviewed for anywhere from one to three years. Secondly, the measured reliabilities of the F-keys were much lower in

TABLE 6.1

LIST OF ITEMS SELECTED FOR NROTC CUMULATIVE EXAMINATION

Item Number on Dec. OCS Exam.	F_1	F_2	F_3
Orient.			
4			1*
7		2	
9		3	
15			4
20	5		
21	6		
24	7		
28			8
40	9		
65	10		
74	11		
75	12		
Ops.			
6			13
12			14
13		15	
20		16	
21			17
26			18
29	19		
67	20		
69	21		
70	22		
71	23		
75	24		
Seam.			
4			25
13			26
18		27	
26	28		
41	29		

Item Number on Dec. OCS Exam.	F_1	F_2	F_3
Nav.			
3			30
4			31
5		32	
7			33
9			34
15		35	
17		36	
20		37	
24	38		
25		39	
29			40
Weap.			
8	41		
11	42		
17	43		
31			44
37			45
38		46	
43			47
47		48	
49		49	
50		50	
52		51	
59		52	
64			53
69		54	
TOTALS (f)	19	17	18

*Numbers in F_1, F_2, and F_3, columns refer to item numbers on cumulative examination.

the NROTC group; indeed, the reliability of Key F-3 in the NROTC group is probably not significantly different from zero. This result is probably due to the fact that the lack of review caused a marked reduction in the inter-correlations of the items. For this group, the items may be said to have been less homogeneous--which items the individual passed was more a matter of chance. In other words, the items were not remembered together; the loss was greater for some items than for others. In contrast, the OCS group had learned and reviewed the items together; they had higher intercorrelations and hence higher reliabilities. These results are a good illustration of the effects of learning and forgetting upon test reliability.

It may also be noted that a possible reason for low reliabilities in the NROTC group may have been the inappropriateness of some of the items for the NROTC curriculum. Perhaps some of the items covered material which had never been taught in the NROTC courses. This possibility, however, is very remote, because the items had been judged by NROTC instructors as appropriate, and several students volunteered the opinion that the Cumulative Examination was perfectly fair--the fault was their own for not having remembered the material.

Since there was a possibility that the two groups were of different ability levels, the analysis of covariance technique was used to control any such differences statistically. Actually, the NROTC group was uniformly superior to the OCS group on OCB subtest scores, as may be seen from Table 6.3; therefore, the analysis of covariance was hardly necessary. Nevertheless, the results shown in Table 6.4, which presents covariance analyses for the three F-scores, will remove all doubt that the differences were significant and in a direction favoring the OCS group, as far as the Cumulative Examination is concerned.

Table 6.5 presents the intercorrelation matrix of F-scores and OCB Sub-test Scores for the NROTC group. These data were utilized in conjunction with data shown in Table 4.8 and elsewhere in order to perform the analysis-of-covariance computations

TABLE 6.2

MEANS, STANDARD DEVIATIONS, AND RELIABILITY AND HOMOGENEITY

DATA FOR THE F-KEYS, CUMULATIVE EXAMINATION

OCS Sample: 200 cases in Class VIII used for iterative factor-analysis of examinations.

NROTC Sample: 156 cases at Harvard, Yale, Brown and Holy Cross administered the Cumulative Examination.

Key		SAMPLE	N	(Rights) $\bar{X}$	σ	K - R (20) rel.	rel (1 item)	H_t
F1	19	OCS	200	14.52	2.0590	.4197	.0367	.0768
F1	19	NROTC	156	12.19	2.1177	.2779	.0198	.0381
F2	17	OCS	200	12.81	2.2827	.5110	.0579	.0956
F2	17	NROTC	156	10.47	2.0707	.2488	.0191	.0314
F3	18	OCS	200	12.32	2.5725	.5540	.0646	.1141
F3	18	NROTC	156	10.30	1.9461	.0690	.0041	.0070

TABLE 6.3

COMPARATIVE DATA ON OFFICER CANDIDATE BATTERY SUBTEST SCORES

OCS Sample: 200 cases in Class VIII used for iterative factor-analysis of examinations.

NROTC Sample: 156 cases at Harvard, Yale, Brown and Holy Cross administered the Cumulative Examination

OCB Subtest	OCS $\bar{X}$	Sample	NROTC $\bar{X}$	Sample
Verbal Reasoning	55.27	8.4486	62.22	7.2178
Mechanical Comprehension	53.42	10.5092	59.12	8.8603
Mathematical Comprehension	50.74	8.5025	59.28	8.7057
Relative Motion	45.88	7.7183	52.02	9.6626
Spatial Relations	53.79	9.3714	57.37	8.9164

presented in Table 6.4. It may be noted that the analysis of covariance performed here was somewhat unusual in that it utilized five control variables (i.e. the five OCB Subtest Scores). Special matrix computational procedures were devised in order to accomplish this.

Comparison of OCS and NROTC students on items freshly reviewed

It was established in the preceding paragraphs that 4th year NROTC students do not retain the material of the first three years very well. True, some of them do, but on the average they do not. The question may be asked, however, what the results would be if OCS students are compared with NROTC students on material which they have both freshly reviewed. That is, how do OCS students compare with NROTC students of the 4th year on material covered in the fourth year of the NROTC curriculum? How do OCS students compare with NROTC students on material covered in the third year, and so on?

The procedure used in answering these questions was as follows:

(1) All items common to the May, 1952 NROTC examinations (by year) and the December 1952 OCS examinations (by department) were identified. Small variations in items, such as an editorial change in the lead, and even a few major variations such as the substitution of an alternative, were ignored in identifying common items. The items were classified in terms of whether they appeared on any of the E-keys (E100, E200, E300 and E400). The results of this analysis are shown in Table 6.6.

(2) It had been hoped that the comparison could be made separately for each of the E-keys, and for each of the four years. This proved impossible in view of the fact that not enough common items were available in each E-key and in each year. It would have been possible, of course, to use all items common to the Examinations, regardless of their appearance in E-keys, but the resulting sets of items would have been heterogenous, and many of them would be of insufficient validity.

TABLE 6.4

COVARIANCE ANALYSES FOR THREE F-KEY SCORES TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN OCS AND NROTC GROUPS AFTER ADJUSTMENT IS MADE FOR OCB DIFFERENCES

Source	S.S.	$(1-R^2)$	Adjusted S.S.	d.f.	Mean Square	F
Key F1:						
Between groups	--	--	484.61	1	484.61	117.1*
Within groups	1547.49	.9337	1444.89	349	4.14	
TOTAL	2027.00	.9516	1929.50	350		
Key F2:						
Between groups	--	--	761.80	1	761.80	198.9*
Within groups	1711.06	.7822	1338.39	349	3.83	
TOTAL	2190.89	.9586	2100.19	350		
Key F3:						
Between groups	--	--	645.47	1	645.47	154.1*
Within groups	1914.36	.7630	1460.66	349	4.19	
TOTAL	2271.49	.9272	2106.13	350		

*Significant far beyond the .001 level.

TABLE 6.5

INTERCORRELATIONS OF THE OFFICER CANDIDATE BATTERY AND "F" SCORES

NROTC Cumulative Examination Sample (N = 156 students at

Harvard, Holy Cross, Brown, Yale)

Variable	1	2	3	4	5	6	7	8
1 OCB (Verbal Reasoning)	1.00	.388	.338	.152	.264	.381	.315	.345
2 OCB (Mechanical Comp.)	.388	1.00	.410	.322	.500	.213	.339	.246
3 OCB (Mathematical Comp.[2])	.338	.410	1.00	.375	.442	.126	.237	.154
4 OCB (Relative Motion)	.152	.322	.375	1.00	.352	.086	.270	.146
5 OCB (Spatial Relations)	.264	.500	.442	.352	1.00	.112	.227	.175
6 F 1 Key	.381	.213	.126	.086	.112	1.00	.278	.207
7 F 2 Key	.315	.339	.237	.270	.227	.278	1.00	.337
8 F 3 Key	.345	.246	.154	.146	.175	.207	.337	1.00

TABLE 6.6

ITEMS WHICH ARE FOUND IN BOTH THE DEC. 1952 OCS EXAMS AND IN MAY 1952 NROTC EXAMS, WITH THE CORRESPONDING ITEM-NUMBERS IN EACH EXAMINATION

	E100		E200		E300		E400		ITEMS NOT INCLUDED IN E KEYS			
	OCS	NROTC	OCS	NROTC	OCS	NROTC	OCS	NROTC	OCS	NROTC	OCS	NROTC
	S17	138	OP13	146			S38	148	OP46	116	0 15	7
									OP51	117	0 23	39
									OP35	114	0 68	77
1NS 1952									OP39	119	S2	88
									OP41	118	S3	92
									0 29	84	S35	130
									0 10	82	S21	137
									0 14	75	S28	135
									0 7	5	S25	139
									0 27	99		
	OP67	108	OP74	109	W37	140	OP57	123	OP75	110		
	OP69	111	W47	55	W41	148	W15	26	W5	15		
2NS	OP71	112	W58	63	W43	119	W48	90	W6	36		
1952			W69	77	W52	61			W9	114		
			W72	85	OP58	124						
			OP68	113	W45	2						
	S26	40	N16	58	OP21	21	OP24	19	OP33	22	S27	42
	S23	37			OP30	20			S46	6	S24	38
3NS-1	N17	63			N3	56			S48	8	S31	44
1952									S47	11	N2	54
									S50	16	N12	48
									S29	41	N26	75
			N42	12	N58	28			N46	16		
					N34	4			N55	25		
					N53	23						
					N54	24						
3NS-2					N32	2						
1952					N52	22						
					N56	26						
					N59	29						
					N60	30						
					N31	1						
	E20	41	E30	5	E25	7	E5	24	E2	13	E45	104
	0 75	142	E15	11	E4	16			E7	25	0 52	123
	0 65	135	E14	14	E19	40			E16	31	0 54	127
	E51	93	E6	20	0 70	146			E12	34	0 71	148
			E13	27	0 72	147			E9	36	0 73	150
			E18	45					E17	39		
			E22	44					E23	47		
4NS			E50	94					E34	55		
1952			E53	97					E66	64		
			E55	99					E72	67		
			0 46	124					E74	72		
			0 53	117					E75	79		
			E 24	51					E70	61		
			E28	53					E43	86		
			E37	60					E42	106		
			E68	65					E48	89		
			E44	87					E46	88		

KEY: 0 = Orientation E = Engineering S = Seamanship
OP = Operations N = Navigation W = Weapons

It was therefore decided to make comparative studies only for the second, third, and fourth years of NROTC (i.e., excluding the first year), and to use only items appearing on key E200 and key E300. The items from these two keys were combined on the ground that they were essentially equivalent, since keys E200 and E300 had previously been shown to be highly correlated (see Chapter IV). Thus, for comparative purposes three sets of items were established, identified as Key E23-2, E23-3, and E23-4, respectively, for the second, third, and fourth years.

(3) These keys were applied:

(a) To 400 cases from OCS Class VIII (i.e., the factor-analysis and the cross-validation samples combined), and

(b) to representative samples of NROTC students tested in May 1952. Four hundred cases in each year were studied. These samples of papers had been selected by the Bureau of Naval Personnel as representative of the several thousand cases tested at all 52 NROTC units.

The statistical results are shown in Table 6.7. Since OCB scores were not readily available for the NROTC examinees, it was impossible to adjust (by the analysis of covariance technique) for any differences in the ability level of the respective groups. Nevertheless, it should be pointed out that there is little reason to fear serious bias in the samples, since each was a random sample from its respective group.

For the 4th and 3rd year comparisons, the differences in means are significant far beyond the .001 level, and favor the NROTC group. The difference in the 2nd-year comparison also favors the NROTC group, but was not significant at the .05 level. It is impossible to assign causes for these differences, in the absence of other information. The differences may be due to any one, or several, of the following conceivable conditions: (a) the NROTC group may be of a higher average ability level; (b) the NROTC group may be able to learn better because of longer periods of time available, in contrast to the 16-weeks course at OCS where all subjects are presented in sixteen

TABLE 6.7

COMPARISON OF OCS AND NROTC GROUPS ON E23-KEYS

N = 400 for each group

Year	Group	Key	$\bar{X}$ (Rights)	σ	C.R.	K-R (20) rel.	rel. (1 item)	H_t
4	NROTC	E23-4	15.99	3.23	7.78*	.650	.078	.116
	OCS	(22 items)	14.24	3.14		.585	.060	.095
3	NROTC	E23-3	10.38	2.42	7.03*	.612	.095	.184
	OCS	(15 items)	9.18	2.42		.624	.100	.214
2	NROTC	E23-2	8.86	1.98	1.64	.505	.078	.118
	OCS	(12 items)	8.64	1.86		.390	.051	.082

*P< .001

By computation from the frequency distributions of scores, also given in the table, it would then appear that (at least in the samples used for this analysis) 314, or 78.5% of the NROTC group attained the critical score of 14, while only 239, or 59.75% of the OCS group attained this score.

Summary

From the standpoint of test results from representative samples, it would seem that the NROTC graduate is a slightly better product, on the average, than the OCS graduate, at least when he is compared with respect to material he has freshly reviewed. The fourth-year NROTC student does not, however, retain the material presented in the first three years as well as might be desired. On the other hand, he may be in a better position to relearn it when this knowledge becomes of use to him in his duties in the fleet.

TABLE 6.8

OPERATING CHARACTERISTIC DATA

Key E23-2
N = 800 (400 2nd year NROTC cases tested May, 1952;
400 OCS cases tested Dec., 1952)

Item Group	Mean Normal deviate value	Proportion Passing — Score Groups (Range and Mean) (Rights): 3-5 / 4.5	6	7	8	9	10	11	12
A	-.12	.299	.371	.491	.517	.624	.725	.842	1.000
B	-.65	.366	.457	.587	.662	.759	.880	.939	1.000
C	-.84	.484	.597	.625	.774	.828	.876	.956	1.000
D	-.99	.336	.677	.700	.812	.877	.913	.974	1.000
Cases in score group:	NROTC	23	33	36	64	79	78	56	31
	OCS	28	29	44	69	80	94	42	14
	TOTAL	51	62	80	133	159	172	98	45

SPECIFICATION OF ITEM GROUPS*

ITEM GROUP A: 2NS#	PN	PO	B: 2NS#	PN	PO	C: 2NS#	PN	PO	D: 2NS#	PN	PO
63	.595	.620	61	.745	.762	148	.720	.868	119	.875	.805
113	.845	.200	109	.735	.765	2	.733	.852	140	.915	.723
55	.420	.610	124	.840	.585	77	.795	.797			
85	.635	.330									

* For each item, its number in the 2NS examination is given, as well as its difficulty index in the NROTC group (PN) and in the OCS group (PO).

TABLE 6.9

OPERATING CHARACTERISTIC DATA

Key E23-3
N = 800 (400 3rd year NROTC cases tested May, 1952;
400 OCS cases tested Dec., 1952)

Item Group	Mean normal deviate value		Proportion Passing — Score Groups (Range and Mean) (Rights)								
			1-6 5.3	7	8	9	10	11	12	13	14-15 14.3
A	+.35		.071	.121	.209	.266	.342	.452	.542	.700	.894
B	-.29		.220	.382	.435	.540	.646	.718	.817	.882	.957
C	-.65		.381	.539	.627	.710	.784	.807	.870	.909	.968
D	-1.69		.771	.845	.900	.923	.934	.977	.988	.982	.995
No. of cases in score group:		NROTC	26	20	36	50	69	60	62	41	36
		OCS	58	48	56	57	56	49	46	14	16
		TOTAL	84	68	92	107	125	109	108	55	52

SPECIFICATION OF ITEM GROUPS*

A			B			C			D		
3-NS #	p_N	p_O	3-NS #	p_N	p_O	3-NS #	p_N	p_O	3-NS #	p_N	p_O
29	.505	.418	34	.693	.650	26	.870	.677	22	.983	.982
28	.423	.487	2	.710	.550	1	.740	.707	4	.963	.970
30	.330	.240	31	.733	.502	33	.507	.868	32	.947	.928
24	.377	.195	23	.813	.227				12	.850	.932

*For each item, its number in the 3-NS examination is given, as well as its difficulty index in the NROTC group (p_N) and in the OCS group (p_O).

TABLE 6.10

OPERATING CHARACTERISTIC DATA

Key E23-4
N = 800 (400 4th year NROTC cases tested May, 1952;
400 OCS cases tested Dec. 1952)

Item Group	Mean normal deviate value	5-10 8.9	11-12 11.6	13	14	15	16	17	18	19	20	21-22 21.4
		Proportion Passing — Score Groups (Range and Mean) (Rights)										
A	+.18	.151	.250	.312	.335	.424	.443	.569	.593	.671	.803	.940
B	-.34	.286	.412	.488	.531	.618	.724	.727	.781	.876	.913	.983
C	-.59	.433	.542	.647	.710	.719	.742	.818	.866	.881	.924	.959
D	-.92	.570	.688	.733	.794	.833	.854	.873	.936	.949	.946	.974
E	-1.28	.714	.859	.884	.910	.889	.920	.926	.960	.951	.976	1.000
No. of cases in score group:	NROTC	25	31	30	36	38	48	62	33	40	33	24
	OCS	53	66	42	49	34	48	46	34	14	9	5
	TOTAL	78	97	72	85	72	96	108	67	54	42	29

SPECIFICATION OF ITEM GROUPS*

A 4-NS #	A p_N	A p_O	B 4-NS #	B p_N	B p_O	C 4-NS #	C p_N	C p_O	D 4-NS #	D p_N	D p_O	E 4-NS #	E p_N	E p_O
20	.455	.302	97	.618	.505	147	.725	.660	27	.835	.727	117	.865	.902
53	.578	.410	44	.513	.677	60	.748	.627	94	.835	.785	99	.905	.882
124	.468	.440	14	.710	.570	45	.725	.745	5	.852	.822	65	.898	.932
87	.555	.377	16	.758	.530	40	.755	.735	11	.880	.805			
			7	.790	.522	51	.753	.745						
			5	.770	.562									

*For each item, its number in the 4-NS examination is given, as well as its difficulty index in the NROTC group (p_N) and in the OCS group (p_O).

CHAPTER VII

SUMMARY AND CONCLUSIONS

This project had as its objective the construction of a new series of examinations for measuring the "integrated" knowledge about naval science demonstrated by graduates of naval officer candidate courses.

The definition and description of what "integrated" knowledge would consist of, and how it could be measured, constituted the major technical problem arising in this research. Two approaches to this problem were explored:

1. An attempt was made to define "integrated" knowledge by logical and rational analysis.
2. An attempt was made, by means of special statistical analyses of test results, to isolate several linearly independent factors of achievement. It was hoped that one or more of these factors would appear to correspond to the notion of "integrated knowledge" developed in the rational analysis of this problem, and that these factors would appear to be distinguishable from other factors which in turn would correspond to the notion of "non-integrated," "mere factual" knowledge.

In the course of the project, various new examination materials (items) were developed, both for use at the Officer Candidate School (Newport, R. I.) and at the 52 NROTC units throughout the country. The logical analysis of the problem of "integrated" knowledge was used as a guide in constructing these examinations, and samples of completed answer sheets resulting from the administration of the examinations were used as the raw data for the statistical analysis.

Logical analysis appeared to disclose three possible ways of defining "integrated knowledge," at least in terms of the types of test items which might be used in testing such knowledge: (a) items testing the ability to utilize facts

from different branches of subject-matter; (b) items depicting realistic situations drawn from job-analyses of shipboard duties of junior naval officers; and (c) items which would require not only the knowledge of relevant facts but also the ability to weigh various elements in the situation, thus testing decision-making and problem-solving functions. To some extent, the definitions are not mutually exclusive, and all three definitions were used in constructing "integrated" items, although definition (c) was accepted as most generally useful and practicable. Attempts were therefore made to construct items which presented realistic and relatively complex problem-situations involving a number of basic knowledges and requiring the evaluation of various elements in these situations. One particularly promising technique was the construction of series of situationally related items; in each series, some items could not be answered by themselves but required information presented or implied in previous items. It was hoped that such series of items would measure the ability to perceive the various elements involved in a situation. Such an ability would presumably be still another facet of "integrated knowledge."

In practice, it was found that certain branches of subject matter lent themselves to the construction of "integrated" items better than other branches. For example, the subject matter known as Naval Operations lends iteself to items of the problem-solving and decision-making type better than the subject matter of Naval History, to take a rather obvious example. This is not to deny, of course, that such a subject matter as Naval History can be tested from the standpoint of "integrated" knowledge and understanding, but it did not lend itself to the approach adopted in the present research.

The success of the project in constructing "integrated examinations" is chiefly a matter of subjective opinion, since there was not objective criterion available for ascertaining the extent of this success. In the last analysis, the results can be judged only by a subjective evaluation of the examinations themselves.

The statistical analysis of test results attempted, however, to see whether different varieties of achievement could be identified. The Wherry-Gaylord iterative factor-analysis procedure was applied to the examinations administered to the December, 1952 graduating class at the Officer Candidate School. Of the four interpretable dimensions of achievement identified in this manner, two were so highly correlated and so nearly alike in appearance that they were considered as one. This one dimension contained the largest number of items and was interpreted as "achievement in problems requiring spatio-temporal reasoning." It was also the dimension best predicted by the subtests of the Officer Candidate Battery. The other two dimensions were tentatively interpreted, respectively, as "achievement in problem requiring verbal reasoning and judgment" and "achievement in test items requiring chiefly recall of isolated facts." Thus, at least two linearly independent dimensions of achievement seem to correspond to the definition of an "integrated" examination as one involving problem-solving and decision-making, and a third dimension of achievement seems to correspond to the notion of a "non-integrated" examination as one which merely involves the recall of isolated facts. It should be said, however, that these dimensions seem to be rather highly inter-correlated, even though presumably linearly independent in the sense customarily employed in factor-analytic theory. That is, officer candidates <u>tend</u> to perform at about the same level on all three dimensions of achievement. It is entirely possible that the three dimensions isolated arose chiefly through errors of sampling; this possibility could be checked in future investigation. Furthermore, the subjective interpretations which have been made for the factors are possibly open to argument. In any case, it was concluded that there is a very strong "general" factor of achievement in officer candidate programs. In other words, there is one major dimension of achievement, and this dimension is measured almost equally well by "integrated" items of the type specially constructed in this project and by "non-integrated" items which chiefly demand recall of factual information.

This is not to say that the effort to construct examinations of an "integrated" type (in contrast to "non-integrated" examinations) is futile. On the contrary, it is probably desirable to emphasize problem-solving and decision-making in the items of an examination battery for officer candidates, even though students perform about equally well on either type. Educators have consistently held to the view that examinations should reflect the objectives of the curriculum. It is believed that naval training authorities intend that their officer candidate curricula seek (among other things) to teach the ability to solve the complex problems arising in fleet operations and to make the sound decisions called for in these operations. This being so, naval science examinations should as far as possible present problems and situations calling for decision-making similar to those which might be encountered in fleet operations.

Educators have also had the experience that examinations set by external or higher authority tend to have an influence on what is emphasized in a curriculum and how it is taught. This influence can be made beneficial if the examinations reflect the intentions of these authorities. Thus, there is every reason to recommend that continued efforts be made to assure that the examinations prepared for naval officer candidate programs emphasize those types of achievement believed most desirable, regardless of the fact that students appear to perform at about the same level on various types of examination materials.

One of the relatively novel features of the present research was the application of a test-analysis technique called the "technique of individual operating characteristics." This technique is designed to disclose the meaning of a test score in terms of the level of performance which it implies on test items. Thus, instead of interpreting a test score in terms of its relative position in a group of scores (as is done with the usual percentile and standard score techniques), one is enabled to interpret it in terms of what item difficulty level the individual is able to

master with some specified probability. By the inspection of items at this difficulty level, one may therefore judge the satisfactoriness of the performance represented by a given score. It is believed that the data assembled by means of this technique and presented in this report will enable naval training authorities to evaluate the achievement of naval officer candidates more realistically than could be done by means of previously available techniques.

A final problem to which this research sought answers was the comparison of officer candidates in the several officer candidate programs. In this research, data were accumulated and analyzed in such a way as to display the differences, if any, between graduates of the NROTC program and the OCS curriculum at Newport, R.I. The analysis was based upon those examination items common to the two curricula and to the corresponding examinations. These comparisons had to be made separately for each year of the NROTC curriculum, which, of course, is spread out over a four-year period rather than being concentrated in an 18-week course as it is at the Officer Candidate School. In this series of comparisons, based on representative samples of NROTC and OCS students, the average NROTC student was consistently superior to the average OCS student. (There was, of course, considerable overlap in the distributions, however.) The superiority of the average NROTC student was increasingly evident from the second to the fourth year of the NROTC curriculum; the most probable explanation of this result is that NROTC students are increasingly highly selected as groups pass through the four years of NROTC program.

These comparisons of NROTC and OCS students were based on examination materials freshly reviewed by each group; for example, OCS students were compared with fourth-year NROTC students only on examination materials found in the fourth-year NROTC curriculum. It was of interest, however, to ascertain the extent to which fourth-year students retain the material learned in the earlier years of the curriculum. For this purpose, a special examination covering first-, second-, and

third-year NROTC material also appearing in the OCS curriculum was prepared and administered to a selected sample of fourth-year NROTC students in four units in the New England area. The scores of these students were compared with the scores of OCS students on the same items; scores on the Officer Candidate Battery were used as a basis for statistical control of ability level. Use of the analysis of covariance technique revealed that the OCS students were very significantly superior to the NROTC students on this special examination. This result, of course, was not unexpected, since the OCS students had freshly reviewed the material while the NROTC students had not. It does point up the fact, however, that NROTC students in their fourth year do not retain the material taught in earlier years to the extent that might be desired. In view of the fact that there is no formal provision for cumulative review in the NROTC curriculum, naval authorities may wish to consider the implications of these results for a possible modification in the plan of the NROTC curriculum.

As a final word of summary, it may be stated that the examinations which have been in use in the NROTC and OCS programs of the past several years are, in the judgment of the present investigators, reasonably reliable and valid examinations. They have been well planned, they represent adequate samplings from the respective curricula, and they function well in discriminating officer candidates of high and low achievement. The present research has explored some ways in which the examinations could be further improved, and some progress in actually improving the examinations may be claimed.

Several recommendations for further research may be made. First, in the construction of new examination materials, the approach utilized in the present project could be further explored. Items identified in this report as appearing to measure problem-solving and decision-making ability could be used as models for the construction of new examination items. This approach could be extended into certain

areas of the naval science curriculum which were not thoroughly worked on in the present project, such as Engineering and Damage Control, and Military Justice. Descriptions of realistic situations which would provide the necessary context for the construction of "integrated" items could be further developed. The preparation of series of inter-related items based on a single situation appears to be a particularly promising technique.

In the meantime, the special factor keys developed in the present research could be validated against on-the-job shipboard performance. This would require tracing the present assignments of former graduates. The factor keys could also be used as tentative criteria for new items which might be constructed; if this plan were to be followed, the items composing these factor keys should continue to be included in future examinations. Concurrently, further research could be undertaken to examine the factorial composition of naval science examinations in order to see whether further evidence could be found for differentiating factual-knowledge examinations from what have been described as "integrated" examinations in the present report.